

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MIEUX COMPRENDRE LA PRODUCTION D'UN JUGEMENT SATISFAISANT
EN ÉVALUATION DE PROGRAMME

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉDUCATION

PAR
AMÉLIE BOSSIROY

MAI 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

De nombreuses personnes ont contribué de façon importante à la réalisation de ce mémoire et j'aimerais les remercier chaleureusement pour leurs conseils et encouragements.

Mes premiers remerciements vont à madame Marthe Hurteau, directrice de ce mémoire et du projet de recherche à l'intérieur duquel celui-ci s'inscrit ainsi que professeure au département d'éducation et pédagogie à l'Université du Québec à Montréal. Je lui exprime ma profonde gratitude pour sa patience, son appui constant et la rapidité avec laquelle elle m'a fait parvenir ses commentaires. L'encadrement qu'elle a su me donner tout au long de la production de cette recherche ainsi que sa très grande connaissance du sujet m'ont été d'une aide inestimable.

J'aimerais également souligner en quelques mots la participation des professeurs spécialistes en mesure et évaluation Juan M. Wood et Pascal Ndinga, qui ont accepté, à titre de correcteurs, d'apporter des commentaires constructifs à ce mémoire. Je les remercie pour l'intérêt qu'ils ont manifesté à l'évaluation de mon mémoire. Je suis aussi reconnaissante envers monsieur Sylvain Houle, professeur au département des sciences comptables de l'UQÀM et membre de l'équipe de recherche, pour sa contribution à la validation du questionnaire d'entrevue. Je veux aussi remercier sincèrement les participants qui ont accepté de nous recevoir et de partager avec nous leur expertise. Sans les précieuses informations qu'ils nous ont données, il n'aurait pas été possible de réaliser ce projet.

J'adresse mes derniers mots de remerciement à mes parents et mes amis qui m'ont encouragée et soutenue dans les moments de solitude et qui, chacun à leur façon, ont participé aux différentes étapes menant au dépôt de ce mémoire.

TABLE DES MATIÈRES

LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX.....	vi
RÉSUMÉ.....	vii
INTRODUCTION.....	1
 CHAPITRE I	
PROBLÉMATIQUE.....	3
1.1 La définition de l'évaluation de programme	3
1.2 L'évolution du domaine de l'évaluation de programme.....	4
1.3 L'état actuel de la situation.....	5
1.3.1 Les remises en question de l'évaluation de programme	6
1.3.2 Les études proposant des pistes d'explication aux difficultés rencontrées par l'évaluation de programme	7
1.4 Les derniers développements en lien avec le raisonnement évaluatif	8
1.5 La question de recherche	10
1.6 Les pertinences	11
1.6.1 La pertinence sociale.....	11
1.6.2 La pertinence scientifique	12
 CHAPITRE II	
CADRE THÉORIQUE.....	13
2.1 Les différentes perspectives en évaluation de programme	13
2.1.1 La perspective de Shadish, Cook et Leviton (1991)	13
2.1.2 La perspective d'Alkin et Christie (2004).....	15
2.1.3 La perspective de Stufflebeam (2000)	16
2.1.4 La perspective de House et Howe (1999)	17
2.1.5 Conclusion sur les perspectives en évaluation de programme	18
2.2 L'impact des développements théoriques sur la pratique.....	19
2.2.1 La nature de l'information requise.....	19

2.2.2 Le processus sous-jacent à la production d'un jugement.....	25
2.2.3 Conclusion sur l'impact des développements théoriques sur la pratique	33
2.3 Synthèse et objectifs de recherche.....	34

CHAPITRE III

DÉMARCHE MÉTHODOLOGIQUE	37
3.1 Le type d'informations requises	37
3.2 Le type de recherche.....	38
3.3 Les choix méthodologiques.....	38
3.3.1 L'échantillon.....	38
3.3.2 La méthode de collecte de données	39
3.4 L'analyse et l'interprétation des données	41
3.5 Les considérations éthiques.....	42
3.6 Les forces et les limites de la recherche	42

CHAPITRE IV

RÉSULTATS	44
4.1 L'analyse des données.....	44
4.1.1 Le codage des données	44
4.1.2 La synthèse des données.....	45
4.2 La présentation des résultats de recherche	47
4.2.1 Objectif 1 : établir ce qu'est un jugement satisfaisant	47
4.2.2 Objectif 2 : identifier les éléments contribuant à générer un jugement satisfaisant.....	51
4.2.3 Objectif 3 : explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisants.....	57
4.2.4 Objectif 4 : explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant.....	65
4.2.5 Synthèse de la présentation des résultats	67

CHAPITRE V

DISCUSSION	70
5.1 Les résultats les plus révélateurs dégagés de l'analyse de données	70

5.2 Les questions qui restent en suspens	74
CONCLUSION.....	76
APPENDICE A QUESTIONNAIRE D'ENTREVUE.....	78
APPENDICE B FORMULAIRE D'INFORMATION ET DE CONSENTEMENT.....	80
RÉFÉRENCES	82

LISTE DES FIGURES

Figure	Page
2.1 L'arbre d'Alkin et Christie regroupant différents modèles d'évaluation sous trois grandes approches. (Tirée d'Alkin et Christie, 2006.)	16
2.2 Proportion de la qualité appréciée et de la qualité mesurée sur laquelle l'évaluateur se base en fonction de la complexité du programme à évaluer. (Tirée Stake et Schwandt, 2006.).....	20
2.3 La double pyramide de Scriven. (Tirée de Scriven, 1990.).....	27
2.4 La modélisation du processus spécifique à l'évaluation de programme: composantes et interrelations (Tirée de Hurteau, Lachappelle et Houle, 2006.)	29
2.5 La démarche sous-jacente au raisonnement évaluatif : l'élaboration du jugement et de son argumentation. (Tirée d'Arens, 2006.)	31
4.1 Synthèse des éléments qui contribuent à générer un jugement satisfaisant	56
4.2 Synthèse des résultats de nos objectifs de recherche sur les éléments et les procédures qui contribuent à générer un jugement satisfaisant aux yeux des clients de l'évaluation.	69

LISTE DES TABLEAUX

Tableau	Page
2.1 Les composantes de l'évaluation et leurs considérations théoriques selon Shadish, Cook et Leviton (1991).....	14
2.2 Les faiblesses majeures de la qualité mesurée et de la qualité appréciée lorsqu'utilisées de façon indépendante	25

RÉSUMÉ

L'évaluation de programme est de plus en plus utilisée pour mesurer la qualité de programmes de tous genres. Paradoxalement, elle fait face à de constantes et nombreuses critiques, notamment sur sa fonction première, soit générer un jugement suffisamment crédible pour être accepté des clients et aider au processus de prise de décisions des activités futures du programme. D'ailleurs, l'étude de Toulemonde (2005) révèle que les clients de l'évaluation utilisent très peu les résultats issus des rapports d'évaluation parce qu'ils ne répondent pas à leurs attentes. Bien que durant les trente dernières années, plusieurs théoriciens ont proposé des processus servant à guider et à structurer la démarche évaluative, ceux-ci ne trouvent pas écho dans la pratique (Stake et *al.*, 1997). Stake et Schwandt (2006) jettent un nouveau regard sur la façon de générer un jugement crédible en introduisant les notions de *qualité mesurée* et de *qualité appréciée*. La présente étude, de type exploratoire, appuie son corpus théorique sur l'apport de ces deux auteurs et tente de recueillir la perception des personnes directement concernées par l'évaluation sur les éléments et les procédures liés à la pratique évaluative qui sont requis pour générer un jugement satisfaisant. Au total, six entrevues individuelles ont été réalisées auprès d'évaluateurs chevronnés et de clients de l'évaluation sélectionnés de façon intentionnelle.

Les résultats de cette étude semblent démontrer qu'un jugement satisfaisant doit s'appuyer sur deux types d'arguments, soit sur des données planifiées, c'est-à-dire recueillies à partir d'une méthodologie bien définie et aussi sur des informations pertinentes qui émergent au cours de l'évaluation. De plus, le jugement doit être utile aux détenteurs d'enjeux, surtout quand vient le temps de prendre des décisions sur les orientations futures du programme. Ensuite, les répondants soutiennent que le jugement doit être explicite pour que les personnes qui le reçoivent le comprennent facilement. Enfin, les résultats de l'étude font aussi ressortir l'importance de solliciter la participation des différents détenteurs d'enjeux tout au long de la démarche évaluative afin de recueillir les diverses perspectives aussi bien au niveau des besoins que des opinions. Ainsi, il semble que les conclusions de notre étude appuient les propos de Stake et Schwandt (2006), à savoir l'urgence que la pratique en évaluation de programme reconnaisse autant l'espace scientifique que l'espace politique dans la formulation d'un jugement. En d'autres termes, une vision plus démocratique de l'évaluation où les détenteurs d'enjeux occupent un rôle important dans le processus évaluatif.

Mots clés :

Jugement évaluatif, évaluation de programme, pratiques évaluatives, raisonnement évaluatif

INTRODUCTION

L'évaluation de programme est un domaine très prolifique, tant sur le plan théorique que sur le plan de la pratique. De nouvelles théories sont continuellement proposées pour faire avancer les connaissances et tenter qu'elle soit un jour reconnue comme une discipline en soi (Dubois et Marceau, 2005). L'évaluation de programme est aussi de plus en plus populaire à l'intérieur des diverses organisations chargées de la gestion des programmes éducatifs et sociaux. Toutefois, malgré tous ces engouements, il n'en demeure pas moins qu'elle fait l'objet de critiques récurrentes concernant sa capacité à remplir adéquatement son mandat, soit générer un jugement suffisamment crédible pour être accepté des clients et aider au processus de prise de décisions des activités futures du programme. Ces constats amènent plusieurs chercheurs à se pencher sur les raisons de cette problématique, ainsi que sur les pistes de solution pour aider les évaluateurs à étoffer et à justifier leur jugement de façon satisfaisante aux yeux des clients de l'évaluation. Néanmoins, il reste encore beaucoup de chemin à faire.

C'est dans cette perspective que la présente recherche vise à mieux comprendre le processus menant à la production d'un jugement satisfaisant aux yeux des clients de l'évaluation de programme, raison d'être de toute démarche évaluative. Cette recherche s'inscrit dans le cadre du troisième volet des activités du programme de recherche de Hurteau et Houle amorcé en 2005 portant sur l'étude des pratiques évaluatives. L'ensemble de cette démarche vise à enrichir le corpus de connaissances théoriques qui entourent le domaine de l'évaluation de programme.

Ainsi, le présent document présente l'ensemble de la démarche scientifique effectuée au volet III du programme de recherche. Le premier chapitre de ce mémoire établit la pertinence de cette étude en décrivant la problématique à laquelle ce domaine est confronté. Au deuxième chapitre est tracé un portrait des développements théoriques du domaine afin d'apporter des éléments de réponse à la question de recherche énoncée dans la problématique. C'est au troisième chapitre que le détail de la méthodologie, qui vise la vérification des

objectifs de recherche énoncés dans le cadre théorique, est abordé. Finalement, les résultats de notre recherche sont présentés au quatrième chapitre, puis font l'objet d'une discussion au cours du dernier et cinquième chapitre.

CHAPITRE I

PROBLÉMATIQUE

Ce premier chapitre est consacré à établir la problématique sur laquelle se penche cette recherche. Pour ce faire, l'état actuel du domaine de l'évaluation de programme y est décrit, incluant le malaise auquel cette activité est confrontée. Puis, les récents développements de la recherche sont exposés. Enfin, le chapitre se termine en présentant la question de recherche de cette étude ainsi que sa pertinence sociale et scientifique.

1.1 La définition de l'évaluation de programme

L'évaluation de programme consiste en une cueillette systématique d'informations sur les différentes composantes d'un programme et sur leurs interactions dans le but d'en faire une description détaillée, d'en comprendre le fonctionnement, et ultimement, de porter un jugement (Guba et Lincoln, 1981; Scriven, 1991; Patton, 1997; Chen, 2005; Hurteau et Houle, 2005a; Mathison, 2005; Nevo, 2006). À titre d'exemple, il peut s'agir d'informations portant sur la clientèle ciblée par le programme, les objectifs poursuivis, les ressources requises, les services offerts, les impacts ou les effets générés (Chen, 2005; Hurteau et Houle, 2006).

La fonction première de l'évaluation est d'émettre un jugement sur la qualité du programme ou d'une de ses composantes en s'appuyant sur une argumentation solide et valide de sorte que les détenteurs d'enjeux¹ de l'évaluation (*stakeholders*) soient convaincus de son bien-fondé. Par la suite, des recommandations, découlant directement du jugement,

¹ De façon générale, Mark, Green et Shaw (2006) considèrent que les détenteurs d'enjeux constituent l'ensemble des personnes qui sont interpellées, à des degrés divers, par les résultats de l'évaluation du programme. Plus précisément, Weiss (1983) propose quatre catégories, toujours valables à ce jour, soit : 1) les preneurs de décisions, 2) les gestionnaires du programme, 3) les intervenants au sein du programme, et 4) les clients, citoyens et membres de groupes.

sont formulées pour aider à éclairer la prise de décisions quant aux orientations futures du programme (House, 1980; Fournier et Smith, 1993; Weiss, 1998; Mathison, 2005).

Si à ses débuts, l'évaluation de programme s'inscrivait dans une perspective qui s'apparentait à la recherche scientifique, elle a vite pris une orientation de reddition de compte. À l'heure actuelle, elle établit l'efficacité des programmes (Borys et *al.*, 2005), c'est-à-dire qu'elle vérifie dans quelle mesure les programmes à l'étude s'acquittent de leurs responsabilités, soit améliorer le bien-être de la clientèle qu'ils visent en offrant, par exemple, des services spécialisés (Nadeau, 1988; Weiss, 1998; Chen, 2005). L'évaluation peut également être conduite dans une perspective de gestion (Borys et *al.*, 2005; Hurteau et Houle, 2005a). À ce moment, son objectif est d'aider un programme à mieux performer en identifiant ses éventuelles failles. Dans tous les cas, l'évaluation de programme débouche sur un jugement, lequel constitue sa raison d'être.

1.2 L'évolution du domaine de l'évaluation de programme

Avant d'aborder le sujet même de cette recherche, il apparaît opportun de situer le développement de l'évaluation de programme dans une perspective d'évolution du domaine afin de mieux comprendre son positionnement actuel.

L'évaluation de programme, à titre de champ professionnel, est un domaine relativement récent, bien que sa pratique soit ancienne (Scriven, 1991; Mathison, 2005). En effet, malgré le fait que ses origines remontent à l'Antiquité (Guba et Lincoln, 1989), ce n'est que depuis la fin des années 1960 que la communauté scientifique la reconnaît comme une activité spécifique et distincte de la recherche. Cette situation est grandement attribuable à l'expansion des programmes sociaux de grande envergure qui a suivi la Seconde Guerre mondiale et la non-efficacité des approches pour les évaluer (Suchman, 1967; Weiss, 1983). En effet, à l'époque, les approches les plus usuelles en évaluation pour juger de la qualité d'un programme étaient étroitement associées à celles auxquelles la recherche scientifique avait recours (Levin-Rozalis, 2003). Toutefois, les praticiens se sont rapidement rendu

compte qu'elles n'étaient pas appropriées pour effectuer de l'évaluation (Suchman, 1967). À titre d'exemple, les évaluations des mégas projets éducatifs *Head Start* et *Follow Through* se sont avérés des échecs magistraux. Cette problématique a amené les théoriciens à concevoir des modèles et des méthodes d'évaluation plus conviviaux qui structurent la pratique évaluative tout en assurant une rigueur à la démarche (Hurteau et Nadeau, 1987; Levin-Rozalis, 2003). Ainsi, les développements, tant théoriques que pratiques, ont permis à l'évaluation de programme de se distinguer de la recherche scientifique et de développer une spécificité qui lui est propre. Elle est aujourd'hui reconnue comme une activité autonome en soi (Madaus et Stufflebeam, 2004; Scriven, 2006). D'ailleurs, les nombreuses publications scientifiques dont elle fait l'objet en font foi.

Bien que l'*évaluation de programme* repose sur des assises et des fondements théoriques provenant d'une variété de domaines d'application bien distincts et bien identifiés, comme l'éducation, la psychologie, la philosophie et la science politique, la *pratique évaluative*, quant à elle, ne bénéficie pas des mêmes avantages et se développe par elle-même (Shadish, Cook et Leviton, 1991 ; Stevenson et Thomas, 2006). Ses fondements se construisent au fur et à mesure du développement du domaine, c'est-à-dire qu'ils reposent sur ce qui émerge des expériences. Cette absence d'ancrage théorique valide et documenté rend la pratique évaluative difficile à encadrer (King, 2003). À ce sujet, Davidson (2005, p.28) mentionne :

« Unlike medicine, evaluation is not a discipline that has been developed by practicing professionals over thousands of years, so we are not yet at the stage where we have huge encyclopaedias that will walk us through any evaluation step-by-step. Even if we did, such "book knowledge" would not be enough. »

1.3 L'état actuel de la situation

La section suivante consiste à faire état de la situation dans laquelle se retrouve l'évaluation de programme actuellement. Puis, nous présenterons les pistes d'explication que certains auteurs proposent en ce qui concerne le malaise retrouvé au sein du domaine.

1.3.1 Les remises en question de l'évaluation de programme

Tel que mentionné précédemment, l'évaluation de programme est de plus en plus utilisée dans les différents milieux pour mesurer la pertinence des programmes. Malgré sa popularité, il n'en demeure pas moins qu'elle fait l'objet de constantes et nombreuses remises en question par ses utilisateurs, principalement sur sa capacité à remplir sa fonction première, soit générer un jugement crédible, ce qui fragilise par le fait même son avenir (Dubois et Marceau, 2005). À ce sujet, l'étude de Toulemonde (2005) révèle que les clients de l'évaluation utilisent très peu les résultats issus des rapports d'évaluation (jugements et recommandations) parce que ces derniers répondent difficilement aux attentes des milieux et ne parviennent pas à contribuer réellement à la prise de décisions.

Certaines études documentent plus précisément la teneur de ces critiques. Ainsi, le Secrétariat du Conseil du Trésor du Canada indique dans une de ses études publiées en 2004 que des 115 rapports d'évaluation qu'ils ont analysés, «32% d'entre eux présentent un raisonnement inadéquat pour appuyer le jugement et que dans 50% des cas, les jugements ne sont pas appuyés par une information pertinente » (Hurteau, Lachapelle et Houle, 2006, p.29). Les résultats de Hurteau, Houle et Mongiat (à paraître) vont dans le même sens. En effet, une métaanalyse réalisée sur 40 rapports d'évaluation de programme leur a permis de conclure que seulement 50% des rapports émettent une « forme quelconque de jugement » et que dans 90% du temps, ce jugement n'est pas satisfaisant parce qu'il ne permet pas de répondre à la question initiale qui avait motivé la démarche évaluative. Leur analyse a également permis de réaliser que plusieurs évaluateurs ont de la difficulté à dépasser le simple stade de la description du programme. Or, si l'évaluation de programme peut recourir à l'enquête pour décrire l'objet d'évaluation (Fournier et Smith, 1993), il faut tout de même qu'elle aille au-delà de cette simple description, c'est-à-dire poser un jugement qui rend compte de la qualité du programme (Davidson, 2005; Hurteau, Lachapelle et Houle, 2006). Ainsi, c'est la notion de jugement qui distingue l'évaluation de programme des autres formes d'enquêtes (Mathison, 2005), lesquelles se limitent à une description.

En résumé, même si on recourt de plus en plus à l'évaluation de programme, les études ci-dessus démontrent qu'il existe bel et bien un problème sur le plan de la qualité des jugements émis à l'intérieur des rapports d'évaluation. Trop souvent, ces derniers sont insuffisamment étoffés et justifiés, ce qui réduit leur crédibilité aux yeux des clients. En conséquence, bon nombre de rapports se retrouvent sur les tablettes, échouant ainsi à leur mandat. Heureusement, de plus en plus d'études se penchent sur ces désastreux résultats et tentent de trouver des pistes d'explication.

1.3.2 Les études proposant des pistes d'explication aux difficultés rencontrées par l'évaluation de programme

En 1972, dans son article intitulé *The Failure of Educational Evaluation*, Guba a abordé le problème retrouvé au sein de l'évaluation de programme sous l'angle des choix stratégiques. Il blâme entre autres le manque d'expertise des évaluateurs à bien identifier et définir les objets et les critères d'évaluation servant à émettre un jugement appuyé par des informations pertinentes et suffisantes, bref à suivre une démarche spécifique d'évaluation de programme. Fournier et Smith (1993), pour leur part, constatent que les évaluateurs accordent trop d'importance à la collecte de données par rapport à la justification de leurs déclarations évaluatives, rendant celles-ci moins crédibles et défendables aux yeux des clients de l'évaluation. En 1995, Scriven arrive à des conclusions similaires à celles émises par Guba en 1972. En effet, son observation de la pratique lui permet de remarquer que les praticiens ne se réfèrent pas aux fondements théoriques en évaluation pour sélectionner les objets et les critères d'évaluation ainsi que pour appuyer leur jugement, ce qui affecte la validité des conclusions évaluatives. Les propos de Dubois et Marceau (2005) ainsi que ceux d'Arens (2006) vont dans le même sens. Ils voient un manque de cohérence entre la théorie et la pratique. Les chercheurs n'utilisent pas les développements réalisés sur le terrain tandis que les praticiens ne se renseignent pas sur les percées dans le domaine de la recherche évaluative. Pourtant, plusieurs chercheurs, tels que Brousselle, Champagne et Contandriopoulos (2006), ont tenté de démontrer qu'il était possible que les théories en évaluation de programme servent à surmonter certaines difficultés survenues dans la pratique.

De leur côté, Virtanen et Uusikylä (2002) arrivent à la conclusion que la pratique actuelle de l'évaluation de programme ne fait qu'émettre des évidences, de façon mécanique, à des problèmes complexes et multidimensionnels de la société moderne. Ces derniers demanderaient plutôt une connaissance approfondie de l'objet d'évaluation afin d'être en mesure de poser des jugements spécifiques. Finalement, plusieurs auteurs, dont Nadeau (1988) ainsi que Donaldson et Lipsey (2006), croient que le manque de formation des évaluateurs serait une des causes du laxisme qui se fait présentement ressentir dans la pratique évaluative.

Somme toute, les motifs invoqués par ces auteurs pour expliquer les difficultés rencontrées par les évaluateurs à émettre un jugement que les clients considéreront satisfaisant font appel à la capacité de raisonner des évaluateurs : difficultés à effectuer les choix stratégiques, à arrimer théorie et praxie et à recueillir les informations nécessaires pour justifier le jugement. Ces réflexions constituent des pistes d'explication plausibles et intéressantes qui ont stimulé différents chercheurs à orienter leurs recherches sur le plan du raisonnement évaluatif.

1.4 Les derniers développements en lien avec le raisonnement évaluatif

Le présent point présente l'état de la question sur les recherches qui ont tenté de documenter la relation entre les difficultés rencontrées par l'évaluation de programme et les explications mises de l'avant, lesquelles réfèrent plus spécifiquement au raisonnement évaluatif. Peu de recherches scientifiques portent sur cette problématique et celles qui existent sont récentes. Alors que les études discutées précédemment relevaient plus du niveau des impressions, les deux recherches qui suivent ont été validées scientifiquement.

Arens (2006) s'est intéressée au processus du raisonnement évaluatif générant un jugement qui s'appuie sur une argumentation. À travers cinq rapports d'évaluation jugés exemplaires par l'*American Journal of Evaluation*, elle a effectué l'analyse de l'argumentation dans le but de décrire et de comprendre de quelle façon le jugement évaluatif

est élaboré. Elle a constaté que les évaluateurs n'explicitent pas, dans leur rapport d'évaluation, la démarche les ayant conduits à poser un jugement et qu'ils ne justifient pas ce dernier au moyen d'une argumentation de qualité. Ces résultats l'amènent à se questionner sur la validité et la crédibilité du jugement et des recommandations qu'émet l'évaluateur à la fin de son rapport. Logiquement, une absence d'arguments valables, vérifiables et exprimés de façon explicite amènera le client à ne pas avoir l'impression que l'évaluateur a suivi une logique de travail et il devra donc se fier à ce que lui dit l'évaluateur. Arens (2006) conclut son étude en mentionnant que le raisonnement évaluatif est un processus intellectuel difficilement identifiable et qu'il serait probablement plus facile de le cerner en discutant directement avec les évaluateurs plutôt qu'en le cherchant à travers des rapports d'évaluation.

L'étude de Hurteau, Houle et Mongiat (à paraître) s'est également intéressée à la production du jugement évaluatif. Tout comme l'étude d'Arens (2006), elle repose sur l'analyse de rapports d'évaluation. Au total, une quarantaine de rapports provenant de la base de données ERIC ont permis aux auteurs de conclure que seulement 50% de ces rapports génèrent une forme quelconque de jugement, ce qui confirme les résultats obtenus par Arens.

Dans leur article *On discerning quality in evaluation*, Stake et Schwandt (2006) jettent un nouveau regard sur le problème de la crédibilité du jugement émis. Pour ces auteurs, le jugement évaluatif devrait s'appuyer sur des informations qui relèvent autant de la légitimité scientifique que des perceptions. Ce constat les amène à introduire les notions de qualité mesurée (*quality as measured*) et de qualité appréciée (*quality as experienced*). Alors que la première réfère aux données quantitatives et qualitatives qui devraient documenter la question d'évaluation à l'origine de la démarche évaluative, contribuant ainsi à la légitimité scientifique, la seconde réfère plutôt aux préoccupations, opinions et attitudes de l'ensemble des détenteurs d'enjeux, et émerge en cours de processus. Ces auteurs concluent que les deux types de qualité devraient toujours être présents dans une démarche évaluative. Van der Maren (1995, p.153) confirme cette position : « Toute évaluation vise la production d'énoncés crédibles qui parviendront à convaincre les lecteurs de leur validité prétendument logique. Pour atteindre cette conviction, l'auteur peut recourir à divers arguments qui relèvent

de stratégies ». Ainsi, pour ces théoriciens, la crédibilité d'un jugement évaluatif s'appuie autant sur la légitimité scientifique que sur les différentes perceptions.

En somme, les conclusions des récents travaux d'Arens (2006) ainsi que ceux de Hurteau, Houle et Mongiat (à paraître) semblent confirmer, dans une certaine mesure, les impressions de Guba (1972) et Scriven (1995) voulant que les remises en question auxquelles fait face l'évaluation de programme puissent être attribuables aux difficultés rencontrées par les évaluateurs à générer un jugement crédible. Quant aux réflexions de Stake et Schwandt (2006), elles apportent deux nouvelles notions, soit la « qualité mesurée » et la « qualité appréciée », lesquelles aideraient l'évaluateur à générer un jugement plus crédible.

1.5 La question de recherche

Bien que les travaux d'Arens (2006) et de Hurteau, Houle et Mongiat (à paraître) décrivent la problématique retrouvée dans les rapports d'évaluation quant à l'élaboration du jugement évaluatif, ils ne fournissent toutefois pas de pistes de solution à cette situation. Par contre, les récents constats de Stake et Schwandt (2006) jettent un nouveau regard. Ainsi, bien que la plupart des recherches scientifiques en évaluation de programme se soient penchées sur la légitimité scientifique (qualité mesurée), aucune d'entre elles n'a encore exploré l'impact que pouvaient avoir les différentes perceptions (qualité appréciée) sur le jugement rendu. Ainsi, il apparaît de plus en plus impératif d'agrandir le spectre des considérations en allant explorer de nouvelles perspectives conceptuelles et procédurales.

La présente étude, qui s'inscrit dans le cadre du troisième volet des activités du programme de recherche de Hurteau et Houle amorcé en 2005 portant sur l'étude des pratiques évaluatives, cherche à développer une meilleure compréhension des mécanismes de production d'un jugement. Plus précisément, elle vise à fournir des éléments de réponse à la question suivante : *De façon générale, quels sont les éléments et les procédures liés à la pratique évaluative qui sont requis pour générer un jugement satisfaisant aux yeux des clients de l'évaluation?*

1.6 Les pertinences

Finalement, nous terminerons le présent chapitre en démontrant la pertinence de cette recherche. D'ailleurs, comme nous l'exposons ci-dessous, plusieurs raisons, d'ordre social ou scientifique, portent à croire qu'il est urgent que davantage de chercheurs se penchent sur le sujet.

1.6.1 La pertinence sociale

Le système d'éducation nord-américain fait souvent appel à des évaluateurs pour évaluer ses différents programmes. Il suffit de penser aux remous provoqués par la récente réforme des bulletins scolaires dans les écoles québécoises. Il faut des arguments de poids pour justifier les changements afin que ceux-ci soient vus comme pertinents aux yeux des clients, soit les parents dans l'exemple précédent. Pourtant, comme il a été mentionné précédemment, l'évaluation de programme parvient difficilement à émettre des jugements convaincants.

Le but ultime de l'évaluation de programme est d'éviter la prise de mauvaises décisions sur un programme donné, laquelle pourrait entraîner la fermeture de programmes importants dont des gens dépendent ou encore le maintien de programmes inutiles (Chen, 2005). Des programmes qui n'apportent pas les effets escomptés ou encore des rapports d'évaluation non utilisés par les clients coûtent très cher. Le développement de ce domaine de recherche est essentiel, car plusieurs programmes sont tributaires des évaluations qui leur sont faites. De plus, mieux comprendre les difficultés et les besoins des évaluateurs aiderait à améliorer la formation de ces derniers et contribuerait à faire pression pour une certification obligatoire.

1.6.2 La pertinence scientifique

L'étude des pratiques en évaluation de programme est également importante sur le plan de la pertinence scientifique. Peu d'études scientifiques traitent spécifiquement de la praxie en

évaluation de programme (Hurteau, Lachapelle et Houle, 2006) et la plupart des débats qui s'y intéressent ne se basent pas sur des données empiriques (Mark, 2008), mais plutôt sur des croyances, témoignages et anecdotes (Smith, 1993; King, 2003). Ce manque flagrant de connaissances empiriques sur les pratiques en évaluation nuit au développement de théories pertinentes et utiles servant à appuyer le jugement évaluatif. Il serait donc important que les théoriciens se penchent davantage sur la pratique évaluative en évaluation de programme afin que des fondements théoriques valables la soutiennent (Shadish, Cook et Leviton, 1991; King, 2003; Christie et Azzam, 2005). À ce sujet, Weiss (2004, p.166) pense que « it is time we spent more time *doing* evaluation, teaching people how to do evaluation, advising on evaluation, reviewing and critiquing evaluation, meta-analyzing evaluations, and in general, advancing the *practice* of evaluation. » Parmi le peu d'auteurs qui s'y intéressent de façon scientifique, la plupart traitent de l'aspect méthodologie, délaissant les enjeux stratégiques, tels que le jugement évaluatif (Arens, 2006).

Considérant l'importance des remises en question et leur impact sur le développement futur de l'évaluation de programme, Mark (2008) souligne qu'il serait urgent que davantage de recherches scientifiques se penchent sur les circonstances dans lesquelles l'évaluation démontre des failles à remplir ses promesses. C'est pour toutes ces raisons que la présente recherche tentera de contribuer au développement de l'état des connaissances dans la pratique évaluative afin qu'un jour ce domaine puisse être reconnu.

CHAPITRE II

CADRE THÉORIQUE

Le présent chapitre est destiné à dresser un portrait général des différentes perspectives qui ont été développées en évaluation de programme, et ce, plutôt en ce qui a trait aux conceptions qu'aux méthodologies. Puis, différents concepts de la pratique évaluative sont développés afin d'apporter des éléments de réponse à la question de recherche énoncée dans la problématique. Finalement, une synthèse des recherches est apportée pour permettre la formulation d'objectifs spécifiques de recherche.

2.1 Les différentes perspectives en évaluation de programme

Dans le chapitre précédent, nous avons vu l'hypothèse soulevée par Scriven (1995), soit que la difficulté rencontrée par les évaluateurs à générer un jugement satisfaisant aux yeux des clients de l'évaluation serait due au fait qu'ils ne se réfèrent pas aux fondements théoriques développés en évaluation de programme. Ceci nous amène à nous attarder à différentes conceptions développées en évaluation de programme à ce jour. Plus précisément, nous allons exposer comment les différents acteurs clés du domaine conçoivent les assises de ce champ d'activité, ce qui nous permettra de mieux cerner notre problématique.

2.1.1 La perspective de Shadish, Cook et Leviton (1991)

Shadish, Cook et Leviton (1991) ont été les premiers à élaborer une métathéorie de l'évaluation de programme. Encore aujourd'hui, celle-ci est considérée, autant par les théoriciens que par les praticiens, comme un classique dans le domaine. De nombreux auteurs, tels que Hurteau et Houle (2005b) ainsi qu'Alkin et Christie (2004), se sont inspirés de la réflexion de Shadish, Cook et Leviton (1991) pour amorcer leurs travaux. Cette métathéorie de l'évaluation de programme comprend cinq composantes (voir tabl. 2.1). La

première est la *théorie de la programmation sociale*. Les auteurs trouvent important de se questionner sur la nature des programmes sociaux, sur la manière dont ils contribuent au changement social et sur la façon dont ils peuvent être améliorés pour mieux remplir leur mission. La seconde, soit la *théorie de la connaissance méthodologique*, s'intéresse aux choix méthodologiques que l'évaluateur doit effectuer selon la situation afin qu'il soit en mesure de recueillir les connaissances pertinentes aidant à la compréhension du problème de départ. La troisième composante est la *théorie de la valeur* et réfère à la manière dont s'élabore et se pose le jugement au cours d'une évaluation. De plus, elle met en évidence le fait que toute démarche évaluative s'inscrit toujours dans un contexte associé à des valeurs, lesquelles doivent être explicitées par l'évaluateur. La quatrième composante, soit la *théorie de l'utilisation*, vise à identifier la façon dont les résultats de l'évaluation seront récupérés dans la pratique, c'est-à-dire par qui, comment et dans quel contexte. Finalement, la *théorie de la pratique*, la dernière composante, fournit des stratégies pour soutenir les praticiens lorsqu'ils ont à faire face à des contraintes dans l'exercice de leur fonction.

Tableau 2.1
Les composantes de l'évaluation et leurs considérations
théoriques selon Shadish, Cook et Leviton (1991)

Composantes théoriques de l'évaluation	Considérations théoriques
• Programmation sociale	source pour planifier et implanter les programmes qui seront évalués
• Connaissance méthodologique	source de connaissance méthodologique
• Valeur	appréciation de la valeur du programme ou de ses composantes
• Utilisation	contexte de procédures qui favorisent l'utilisation des résultats de l'évaluation
• Pratique évaluative	l'art de faire et la connaissance des évaluateurs

Tiré de Stevenson et Thomas (2006)

2.1.2 La perspective d'Alkin et Christie (2004)

Alkin et Christie (2004) poursuivent la réflexion émanant des travaux de Shadish, Cook et Leviton (1991) et élaborent une deuxième métathéorie de l'évaluation de programme. Toutefois, ils décident de s'éloigner des théories traditionnelles en évaluation de programme pour plutôt s'intéresser aux attributs communs qu'il est possible d'identifier dans les différents modèles d'évaluation. Effectivement, au cours des dernières décennies, plusieurs auteurs de référence ont élaboré des modèles d'évaluation dans le but d'offrir un processus pour structurer la démarche évaluative et guider l'évaluateur dans la formulation de jugements défendables (Madaus et Kellaghan, 2000). Comme il existe un grand nombre de modèles en évaluation de programme, Alkin et Christie (2004) en sont venus à proposer une typologie de 28 auteurs et de leurs modèles d'évaluation respectifs, illustrée au moyen d'un arbre (voir fig. 2.1). Ces modèles se différencient entre eux par leur approche théorique. Ils représentent la vision idéalisée d'un auteur sur un concept important (Stufflebeam, 2000; Alkin et Christie, 2004). Ainsi, ils se retrouvent à mi-chemin entre l'idéologie et la méthodologie (Scriven, 2000).

Alkin et Christie (2004) ont donc essayé de regrouper les modèles sous trois grandes approches théoriques, chacune représentée par une branche d'arbre. La branche de la *méthodologie*, inspirée de la construction de la connaissance méthodologique de la métathéorie de Shadish, Cook et Leviton (1991), regroupe les auteurs qui axent leurs efforts sur l'aspect méthodologique d'une évaluation afin d'aller chercher toute l'information pertinente. La branche du *jugement de valeur*, inclut les modèles qui focalisent sur le rôle fondamental de l'évaluateur, soit établir la valeur réelle de l'objet d'évaluation pour éclairer la prise de décision. Les théoriciens identifiés à cette approche « maintain that placing value on data is perhaps the most essential component of the evaluator's work. » (Alkin et Christie, 2004, p.13) Finalement, les théoriciens rassemblés sur la branche de l'*utilisation* sont ceux qui sont constamment concernés par la récupération des résultats émanant de l'évaluation une fois celle-ci terminée. Autrement dit, ils sollicitent la participation des clients de l'évaluation tout au long de la démarche évaluative pour que ceux-ci s'approprient l'évaluation et utilisent leurs recommandations. Alkin et Christie (2004) précisent que ces

regroupements ne sont pas mutuellement exclusifs. Ils ont effectué leur répartition en fonction de l'approche préconisée par les théoriciens, même si bien souvent, les trois approches étaient présentes, à des degrés divers, à l'intérieur de leur modèle.

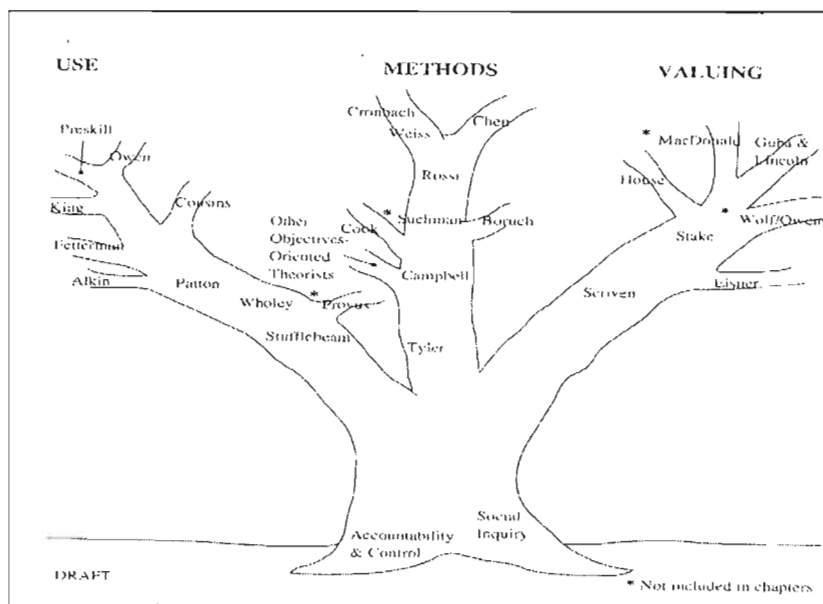


Figure 2.1 L'arbre d'Alkin et Christie regroupant différents modèles d'évaluation sous trois grandes approches. (Tirée d'Alkin et Christie, 2006.)

2.1.3 La perspective de Stufflebeam (2000)

Stufflebeam (2000) relève 22 modèles en évaluation de programme qu'il est possible de regrouper en deux grands courants de pensée. Le premier, caractérisé par les premières générations d'évaluateurs issus du domaine des sciences sociales, est nettement influencé par la reddition de compte. Ce paradigme scientifique se distingue par le recours systématique aux méthodes expérimentales ainsi qu'aux instruments standardisés. Bien que cette orientation soit encore présente dans la pratique, les évaluateurs suivants ont senti le besoin de développer de nouvelles perspectives qui ne visent pas tant l'objectivité de l'information comme la nécessité de répondre aux besoins d'information des clients de l'évaluation, et ce, en les associant au processus. Ce deuxième courant de pensée, motivé par le fait que les

évaluations sont peu utilisées par les clients, s'inscrit donc dans un mouvement de démocratisation qui tente de surmonter les difficultés auxquelles cette activité fait face en la rendant plus inclusive, transparente et démocratique. Plusieurs études ont été effectuées sur les effets de cette nouvelle orientation et si les résultats ne sont pas toujours concluants, il semble se dégager une tendance selon laquelle les évaluations s'avèrent plus crédibles et utiles (Ryan, 1998; Johnson, Willeke et Steine, 1998; Fine, Thayer et Coghlan, 2000; Thayer et Fine, 2000; Unrau, 2001).

2.1.4 La perspective de House et Howe (1999)

House et Howe (1999) proposent des regroupements de courants de pensée qui s'inscrivent dans une perspective de développement du domaine. Étant donné que le présent projet de recherche ne s'intéresse pas au côté historique de l'évolution des perspectives en évaluation de programme, seul le plus actuel des quatre points de vue sera présenté. Il s'agit de la *démocratie délibérative*, vision à laquelle House et Howe adhèrent.

Dans les sociétés démocratiques, le pouvoir décisionnel des activités d'un programme est partagé, à des degrés divers, entre les instances gouvernementales, l'évaluateur et les citoyens. Par conséquent, les évaluateurs sont fortement invités à recourir et à respecter trois principes démocratiques tout au long de leur démarche évaluative afin d'arriver à un consensus entre ces différentes parties soit : l'inclusion, le dialogue et la délibération. L'*inclusion* garantit que toutes les personnes concernées par le programme soient intégrées au processus. Le *dialogue* voit à ce que l'ensemble des points de vue soit représenté et la *délibération*, quant à elle, gère les éventuels conflits qui émergent de la confrontation de valeurs et de l'interprétation des résultats. Quand l'évaluation satisfait à de telles exigences, elle peut être dite démocratique, impartiale et objective. Au contraire, lorsqu'un des principes est absent, il en résulte un déséquilibre des forces et des tensions sont ainsi créées (Abma et al, 2001: voir House et Howe, 1999). House et Howe (1999) estiment qu'avec cette vision, les faits et les jugements de valeur se rejoignent sur un continuum qui permet de déterminer si une affirmation est vraie et objective. Il y a une obligation de fournir au public une

information fondée, honnête et sans biais. House et Howe terminent en soulignant que les approches évaluatives qui s'appuient sur la participation et la collaboration des clients de l'évaluation peuvent augmenter l'efficacité de la collecte de données, améliorer l'évaluation et renforcer l'utilisation des résultats générés par l'évaluation.

2.1.5 Conclusion sur les perspectives en évaluation de programme

Suite aux différentes considérations que nous venons de relever, il semble que les différentes perspectives tirent profit des expériences empiriques, et ce, qu'elles soient positives ou négatives (Shadish, Cook et Leviton, 1991). La pratique de l'évaluation de programme, qui se s'inscrit actuellement dans le courant de la démocratie délibérative, est caractérisée par des fondements théoriques plus pragmatiques dans lesquels :

- Le jugement de valeur fondé et argumenté se trouve au cœur de la démarche évaluative;
- Les données et évidences recueillies sont nuancées par les croyances et les valeurs de l'environnement, lesquelles sont particulières au contexte précis du déroulement de l'évaluation (Schwandt, 2007);
- Les clients de l'évaluation sont impliqués dans toutes les phases du processus. Ainsi, les principes d'inclusion, de dialogue et de délibération sont respectés;
- Aucune loyauté méthodologique n'est entretenue, mais la rigueur scientifique est de mise. Ainsi, l'évaluation de programme a recours sans hésitation à des approches mixtes dont les choix se basent sur le type d'investigation, les données et les analyses requises (Rog et Fournier, 1997). La rigueur scientifique constitue un moyen et non une fin en soi afin d'assurer une crédibilité à la démarche et à son produit.

Ce tour d'horizon des différentes perspectives théoriques qui entourent le concept de l'évaluation de programme est nécessaire pour aider les théoriciens et les praticiens à mieux comprendre la production d'un jugement qui satisfait les clients de l'évaluation. Mieux comprendre cette activité permet de mieux intervenir. Cette idée corrobore l'hypothèse émise

par Arens (2006) selon laquelle les difficultés rencontrées en évaluation de programme seraient dues à un manque de cohérence entre la théorie et la pratique.

2.2 L'impact des développements théoriques sur la pratique

La vision de l'évaluation décrite ci-dessus a des impacts sur notre façon actuelle de concevoir la pratique de l'évaluation de programme. Celle-ci peut être décrite en fonction de deux composantes principales qui se complètent l'une l'autre. La première est *la nature de l'information requise* pour porter un jugement, laquelle réfère aux énoncés servant d'argumentation au jugement contenu de la démarche évaluative. La seconde composante principale de la pratique évaluative se rapporte à la façon de mener l'évaluation, c'est-à-dire le choix du *processus sous-jacent à la production d'un jugement* (les procédures pour encadrer la démarche évaluative). La prochaine section consiste donc à définir plus en profondeur chacune de ces deux composantes principales, lesquelles contribuent à la production d'un jugement, ainsi qu'à présenter ce que les auteurs proposent pour soutenir la pratique.

2.2.1 La nature de l'information requise

Tel que mentionné, l'évaluation de programme consiste essentiellement à porter un jugement sur la qualité d'un programme, mais, trop souvent, ce jugement ne parvient pas à convaincre les lecteurs du rapport. À ce propos, nous avons déjà vu rapidement dans le premier chapitre de ce mémoire que Stake et Schwandt (2006) jettent un nouveau regard sur le problème de la crédibilité du jugement émis. Selon eux, celui-ci devrait s'appuyer sur des informations qui relèvent de deux types de qualité, à savoir la *qualité mesurée* (*quality as measured*) et la *qualité appréciée* (*quality as experienced*). La première établit des données et des évidences qui reposent sur des critères opérationnels ainsi que sur des mesures. Elle contribue ainsi à la légitimité scientifique. La seconde, quant à elle, relève davantage des

perceptions. Elle se fonde sur la connaissance expérientielle (clinique, pratique), sur les préoccupations et sur les significations subjectives des différents détenteurs d'enjeux du programme ainsi que de celles de l'évaluateur lui-même.

Bien que les proportions varient selon les circonstances, l'évaluateur chevronné sait qu'il doit toujours recourir aux deux types de qualité afin d'assurer des assises à son argumentation et ainsi émettre un jugement pertinent qui satisfait l'ensemble des clients de l'évaluation de par sa crédibilité. En effet, si son évaluation repose surtout sur une qualité mesurée, elle sera perçue comme un travail trop mécanique. Inversement, si elle se base uniquement sur la qualité appréciée, elle sera vue comme trop subjective (Stake et Schwandt, 2006). La figure 2.2 représente les proportions de chacune des qualités qu'il est normalement souhaitable de retrouver selon la complexité du programme à évaluer.

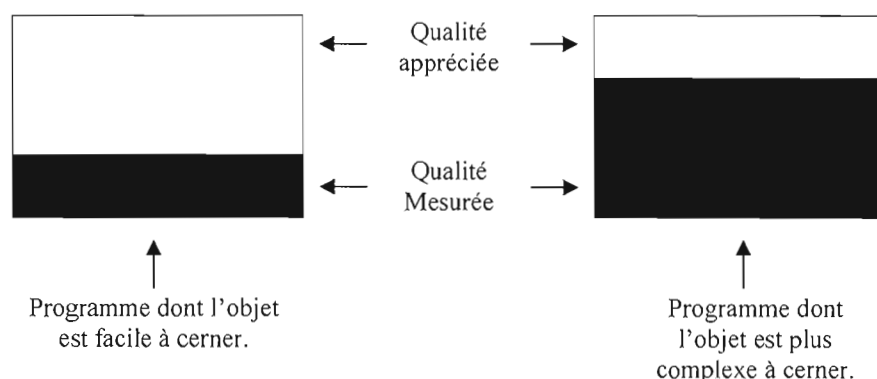


Figure 2.2 Proportion de la qualité appréciée et de la qualité mesurée sur laquelle l'évaluateur se base en fonction de la complexité du programme à évaluer. (Tirée de Stake et Schwandt, 2006.)

Ainsi, tel qu'en fait foi la figure 2.2, dans le cadre d'un programme de petite envergure, l'évaluateur peut fonder son jugement pratiquement uniquement sur une qualité appréciée parce qu'il est en mesure de saisir l'objet d'évaluation dans son ensemble. Par contre, lors de programmes de grande envergure ou faisant l'objet d'application dans de multiples sites, l'évaluateur doit davantage s'appuyer sur la qualité mesurée puisque le programme est alors plus complexe à comprendre (Stake et Schwandt, 2006).

En résumé, bien que la pondération en ce qui concerne les deux types de qualité puisse varier selon la situation, Stake et Schwandt (2006) insistent sur le fait que chacune d'entre elles devait être minimalement présente dans la démarche évaluative pour assurer un jugement crédible. Dans les deux sections ci-dessous, nous allons développer plus en profondeur les deux types de qualité afin de mieux comprendre ce qu'entendent les auteurs par ces notions et comment elles s'établissent.

2.2.1.1 Établir la qualité mesurée d'un programme

À la difficulté de mesurer la bonne chose, s'ajoute celle de bien la mesurer. La qualité est plus difficile à cerner dans les programmes de grande envergure parce que, comme il a été mentionné précédemment, ils sont souvent plus complexes ou nébuleux et couvrent de multiples sites. Généralement, l'évaluateur sera confronté à des composantes du programme qui dépassent ses connaissances personnelles et à des données plus complexes à recueillir. Devant cette réalité, l'évaluateur est souvent obligé d'intervenir de façon indirecte avec les clients de l'évaluation parce qu'ils sont difficilement joignables (Stake et Schwandt, 2006) ainsi que de recourir à des modalités quantitatives pour faciliter son travail, c'est-à-dire à la qualité mesurée.

Dans le cas de la qualité mesurée, l'évaluateur recueille des données et des évidences qui sont constituées d'indicateurs observables et mesurables en fonction des critères préalablement déterminés. Puis, il les compare à des seuils de performance formels et explicites pour ainsi pouvoir juger de la qualité du programme. Ces derniers peuvent être développés par des experts ou être connus d'avance (principes, normes sociales, hypothèses, groupes contrôles, performances précédentes du même programme ou de programmes similaires, etc.) (Scriven 1967; House 1980; Fink, 2004; Stake 2004). Quant aux critères, les auteurs s'entendent pour dire qu'ils doivent être déterminés selon les objectifs de l'évaluation, le contexte particulier dans lequel se situe l'objet d'évaluation, ainsi que selon les besoins et les attentes des personnes directement concernées par l'évaluation (Nevo, 2006). La qualité mesurée aide donc l'évaluateur à rester le plus impartial, rigoureux

et objectif possible. Dans ce sens, elle contribue à la *légitimité scientifique* du jugement. Il faut savoir que cette démarche fait référence à ce qui devrait se passer dans le meilleur des mondes. Par contre, la réalité en est souvent autrement.

Effectivement, Stake et Schwandt (2006) soulignent que les évaluateurs ont tendance à utiliser des instruments de mesure standardisés (ex. : test, questionnaire) lorsqu'ils effectuent une évaluation de programme, que celle-ci soit de petite ou de grande envergure. La plupart du temps, ce choix méthodologique est une solution de facilité, ou un moyen de se rassurer parce que plusieurs croient, à tort, que ces instruments apportent de la rigueur à la démarche scientifique (Stake, 2001). Au contraire, ces tests sont rarement constitués de critères adaptés au contexte spécifique de l'évaluation et d'indicateurs valides, ce qui les amène à émettre des résultats absurdes. À ce propos, Stake (2001, p.350) souligne un problème dans le domaine de l'éducation: « [...] the tests often do not effectively tell what the teachers have taught and the learners have learned. » Sachant que tout programme est unique puisqu'il a lieu dans un environnement précis, à un moment particulier et qu'il implique des personnes qui ont des besoins et des valeurs qui leur sont propres, il est inutile de tenter de trouver le meilleur instrument de mesure, c'est-à-dire celui qui conviendrait à toute évaluation (Scriven, 1972b; House, 1980; Guba et Lincoln, 1981; Stake, 2001; Nevo, 2006). Pour une activité aussi complexe que l'évaluation, il faudrait que l'évaluateur développe des instruments, des outils et des procédures d'évaluation « sur mesure » pour chaque programme (House et Howe, 1999; Stake et Schwandt, 2006).

Néanmoins, il semble que la tâche de certains évaluateurs se résumerait uniquement à appliquer de façon machinale des instruments de mesure standardisés, donc par le fait même souvent inadaptés au contexte spécifique d'évaluation. Par conséquent, Stake (2001) se demande s'il ne serait pas envisageable de relayer le travail des évaluateurs (spécialistes) à des techniciens. Ceci aurait peut-être comme effet d'inciter les évaluateurs à démontrer leurs compétences professionnelles en ne recourant plus seulement à la qualité mesurée, mais en faisant également appel à la qualité appréciée.

2.2.1.2 Établir la qualité appréciée d'un programme

Certaines données échappent à la rigueur de nos méthodes, peu importe à quel point elles sont technologiquement sophistiquées. Ainsi, la qualité appréciée apporte un équilibre. Elle vise à produire un jugement sur les composantes non mesurables de l'objet d'évaluation. Le jugement évaluatif ne se base donc plus sur les données recueillies, mais s'appuie sur des critères implicites, l'expérience, les impressions, les perceptions, l'instinct professionnel et la connaissance pratique de l'ensemble des détenteurs d'enjeux ainsi que de l'évaluateur lui-même (Stake et Schwandt, 2006).

Pour Molander (2002 : voir Stake et Schwandt, 2006), la connaissance pratique englobe autant le savoir que le savoir-faire de l'évaluateur. Ainsi, un manque de connaissance de l'objet d'évaluation ou des pressions exercées sur l'évaluateur par des clients sont des facteurs qui peuvent venir fausser le jugement. Quant à Stake et Schwandt (2006), ils associent la connaissance pratique avec les termes cognitif et affectif. À titre d'exemple, la propre compréhension de l'évaluateur face au phénomène dans l'action, ses valeurs ou ses préjugés personnels peuvent influencer, de façon négative, sa perception de l'objet à évaluer, donc, par le fait même, les résultats de l'évaluation (Hurteau et Nadeau, 1987; Stake et Schwandt, 2006). De même, les perceptions de l'évaluateur peuvent varier selon son rapport avec l'action ainsi que selon la perspective qu'il choisit : celle de l'évaluateur, du lecteur du rapport ou des clients de l'évaluation. Il revient à dire que les perceptions de qualité appréciée sont arbitraires, variables et changeantes (Stake et Schwandt, 2006).

Molander (2002 : voir Stake et Schwandt, 2006) insiste aussi sur l'importance de faire participer les détenteurs d'enjeux au processus de l'évaluation. L'implication des détenteurs d'enjeux serait essentielle pour teinter d'équité la démarche évaluative (House, 1991). L'information apportée par ces personnes aiderait non seulement l'évaluateur à développer une meilleure compréhension du programme, mais favoriserait aussi l'utilisation ultérieure des résultats de l'évaluation (Fetterman, 2001; Patton, 1997). Toutefois, bien que le mouvement démocratique qui valorise l'association des détenteurs d'enjeux à l'ensemble de la démarche évaluative s'est amplifié au courant des dernières années, il semble que cette

pratique ne soit pas encore au point. Toutlemonde (2005) remarque que les évaluateurs éprouvent souvent de la difficulté à prendre en considération les points de vue des différents détenteurs d'enjeux pour construire leur jugement. D'ailleurs, il arrive fréquemment que le jugement évaluatif soit contesté ou réfuté lorsque les détenteurs d'enjeux ont l'impression que leur point de vue n'est pas représenté adéquatement.

Finalement, Stake et Schwandt (2006) reconnaissent que la qualité appréciée n'est pas toujours facile à documenter. Cette démarche génère souvent beaucoup d'informations qui deviennent difficilement gérables si l'évaluateur ne procède pas à un triage afin de dégager les réels enjeux. Encore une fois, cette opération peut créer des tensions entre l'évaluateur et les détenteurs d'enjeux s'ils n'approuvent pas son analyse de la situation. Par ailleurs, Toulmin (1993) soulève un autre point de vue quant à la difficulté éprouvée par les évaluateurs à documenter la qualité appréciée. Selon cet auteur, les informations relevant de la qualité appréciée seraient rarement enracinées dans l'explication ou la justification. Cette approche reposerait plutôt sur l'acuité de la personne ainsi que sur sa crédibilité (Stake et Schwandt, 2006). Il arrive d'entendre que certaines personnes ont le « flair ». Bien souvent, les évaluateurs d'expérience sont capables de tirer des conclusions en un instant, sans passer par aucune des étapes essentielles de la démarche évaluative, c'est-à-dire par un raisonnement laborieux et par une argumentation. Le sérieux avec lequel on envisage leurs propos dépend, par exemple, de la sorte d'homme qu'ils sont (leadership, charisme) et du crédit dont ils jouissent. Ainsi, on se fie au jugement de certains évaluateurs simplement parce qu'ils sont connus ou reconnus pour leur compétence et leur rigueur (Toulmin, 1993).

En somme, bien que la qualité mesurée et la qualité appréciée soient toutes deux importantes, seules, elles ne pourraient pas être totalement efficaces parce qu'elles comporteraient certaines faiblesses (voir tabl. 2.2).

Tableau 2.2
Les faiblesses majeures de la qualité mesurée et de la qualité appréciée
lorsqu'utilisées de façon indépendante

Qualité mesurée	Qualité appréciée
<ul style="list-style-type: none"> • Les critères obtenus scientifiquement, par un expert ou par des clients de l'évaluation, sont rares. • Les seuils de performance sont difficiles à établir, ce qui fait qu'ils ne sont presque jamais présentés dans les rapports d'évaluation (Arens, 2006). 	<ul style="list-style-type: none"> • Approche qui repose uniquement sur l'acuité et la crédibilité de l'évaluateur. • Inexistence de critères adéquats pour mesurer la valeur des témoignages et ainsi contrôler les biais.

À première vue, le point faible de la qualité mesurée serait que l'objet évalué est trop complexe pour qu'il puisse être évalué adéquatement au moyen de quelques critères. Ceci empêche de le saisir dans son ensemble (Stake, 2001). Quant à la qualité appréciée, il s'agirait de son manque de rigueur dû à une absence de méthodologies articulées et complexes. C'est pour cette raison que Stake et Schwandt (2006) insistent sur le fait que les deux qualités sont essentielles, bien que leur proportion varie selon la nature et l'ampleur du programme à l'étude. Utilisées de façon complémentaire, elles contribuent à augmenter le niveau de crédibilité du jugement, le rendant, par le fait même, satisfaisant aux yeux des clients de l'évaluation. Cela dit, il reste qu'il n'est pas évident de synthétiser et d'intégrer les informations issues de chacune des deux formes de qualité pour produire un tel jugement. Cette dernière remarque introduit le point suivant qui porte sur les processus sous-jacents à la production d'un jugement.

2.2.2 Le processus sous-jacent à la production d'un jugement

La deuxième composante principale de la pratique évaluative, complémentaire à la nature de l'information requise, est le processus servant à mener l'évaluation afin d'obtenir un

jugement crédible. Comme nous l'avons déjà exposé dans le chapitre de la problématique, Guba (1972) propose l'hypothèse que les critiques adressées aux évaluateurs en ce qui concerne le jugement crédible émis dans les rapports sont dues à leur difficulté à suivre une démarche spécifique en évaluation de programme. À ce propos, l'évaluation de programme bénéficie de plusieurs processus qui lui sont propres, tous fortement inspirés de la *logique de l'évaluation* de Scriven (1980). Dans le cadre de ce mémoire, nous présenterons deux de ces processus, soit la modélisation de l'acte spécifique d'évaluer d'Hurteau, Lachapelle et Houle (2006) et le raisonnement évaluatif d'Arens (2006). Bien que ceux-ci guident la démarche évaluative, ils ne sont pas un gage d'évaluation de qualité. Comme nous le verrons, leurs composantes constituent plus des caractéristiques pour décrire le processus de l'évaluation que pour déterminer la qualité du programme en question.

Étant donné que Scriven (1980) est l'un des premiers auteurs s'étant intéressé aux fondements de la démarche logique en évaluation de programme et qu'il a grandement contribué au développement théorique de l'acte spécifique d'évaluer, il est nécessaire de s'arrêter ici pour mieux définir sa contribution.

2.2.2.1 La logique de l'évaluation (Scriven, 1980)

Les années 1970 et 1980 ont donné naissance à de nombreux modèles et méthodes visant à structurer la démarche évaluative (Madaus et Stufflebeam, 2004). Toutefois, seul Scriven (1980), en s'investissant dans la recherche d'une définition plus raffinée des composantes qui caractérisent l'acte spécifique d'évaluer, est parvenu à légitimer la démarche. Pour ce faire, il a élaboré une métathéorie de sa propre conception de l'évaluation, soit la *logique de l'évaluation*. Celle-ci consiste en un raisonnement spécifique qui représente les paramètres et les processus fondamentaux de toute évaluation (Scriven, 1995).

Plus spécifiquement, le concept de la *logique de l'évaluation* de Scriven (1980) regroupe quatre opérations, lesquelles se retrouvent à la page suivante.

1. *Établir les critères* pour connaître par rapport à quels éléments ou composantes l'objet d'étude sera évalué. Ces critères sont les aspects, les attributs ou les dimensions qui caractérisent la valeur de l'objet d'évaluation. Ils servent donc de point de référence au jugement ;
2. *Établir les seuils de performance* attendus pour chacun des critères. Ils constituent en quelque sorte le point de coupure entre ce qui constitue un jugement favorable ou défavorable ;
3. *Mesurer la performance* (indicateurs observables et mesurables) et la comparer aux seuils de performance établis afin de savoir dans quelle mesure l'objet a performé. Il s'agit donc de l'analyse des données recueillies;
4. Synthétiser et intégrer les données afin d'être en mesure de *porter un jugement* sur la qualité ou le mérite de l'objet d'étude. C'est l'étape qui se trouve au cœur de la démarche évaluative.

Plus tard, Scriven (1990) poursuit ses réflexions et propose le concept de la *double pyramide* (voir fig. 2.3), lequel complète sa vision de la pratique évaluative. La première pyramide rend compte du processus analytique, processus qui consiste à identifier les critères, les indicateurs et les données. Quant à la deuxième pyramide, soit la pyramide renversée, elle illustre le processus synthétique, lequel consiste à partir des informations ramassées plus tôt pour faire des inférences, des synthèses et ainsi arriver à un jugement.

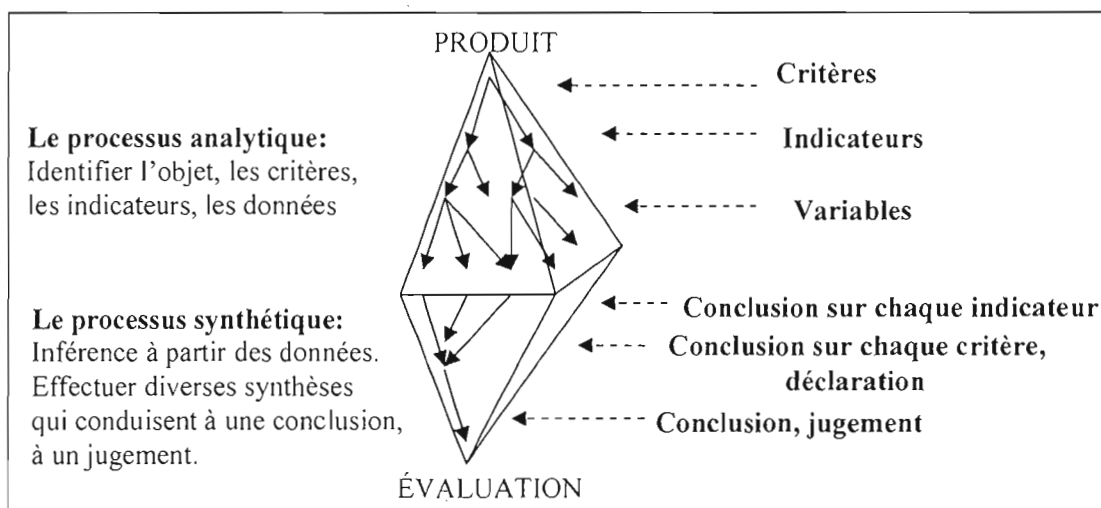


Figure 2.3 La double pyramide de Scriven. (Tirée de Scriven, 1990.)

2.2.2.2 La modélisation de l'acte spécifique d'évaluer (Hurteau, Lachapelle et Houle, 2006)

La diversité des modèles et des approches en évaluation de programme n'a pas seulement intrigué Scriven. Hurteau et Houle (2005b) se sont également intéressés à un phénomène qu'il est possible d'observer dans la pratique, à savoir que l'évaluation semble suivre un ensemble logique d'opérations spécifiques générant un jugement de valeur sur l'objet à évaluer. À ce sujet, un consensus commence à émerger au regard des principes qui devraient rester les mêmes d'une évaluation à l'autre.

Récemment, dans le cadre de leurs travaux portant sur l'analyse des pratiques évaluatives, Hurteau et Houle (2005b) ont fait une anasynthèse (Silvéry, 1972; Legendre, 1993; Landry et Auger, 2003) des différentes contributions théoriques descriptives en évaluation de programme. Ils se sont notamment appuyés sur le référentiel théorique fourni par la *logique de l'évaluation* de Scriven (1980), tel que vu précédemment, ainsi que sur les contributions ultérieures (Hurteau, 1991, Fournier, 1995, Scriven, 1995, Stake, 1995) pour compléter chacune des composantes de leur anasynthèse. Chacune de ces contributions est rapidement développée ci-dessous.

Cette anasynthèse a amené Hurteau et Houle (2005b) à élaborer la modélisation de l'acte d'évaluer, processus spécifique à la pratique de l'évaluation de programme présenté au moyen de la figure 2.4, située à la page suivante. Ce processus générique peut être appliqué pour toute forme d'évaluation. Elle offre aux évaluateurs un cadre de référence qui guide leur pratique et leur permet de générer un jugement crédible, puisque fondé sur des choix stratégiques.

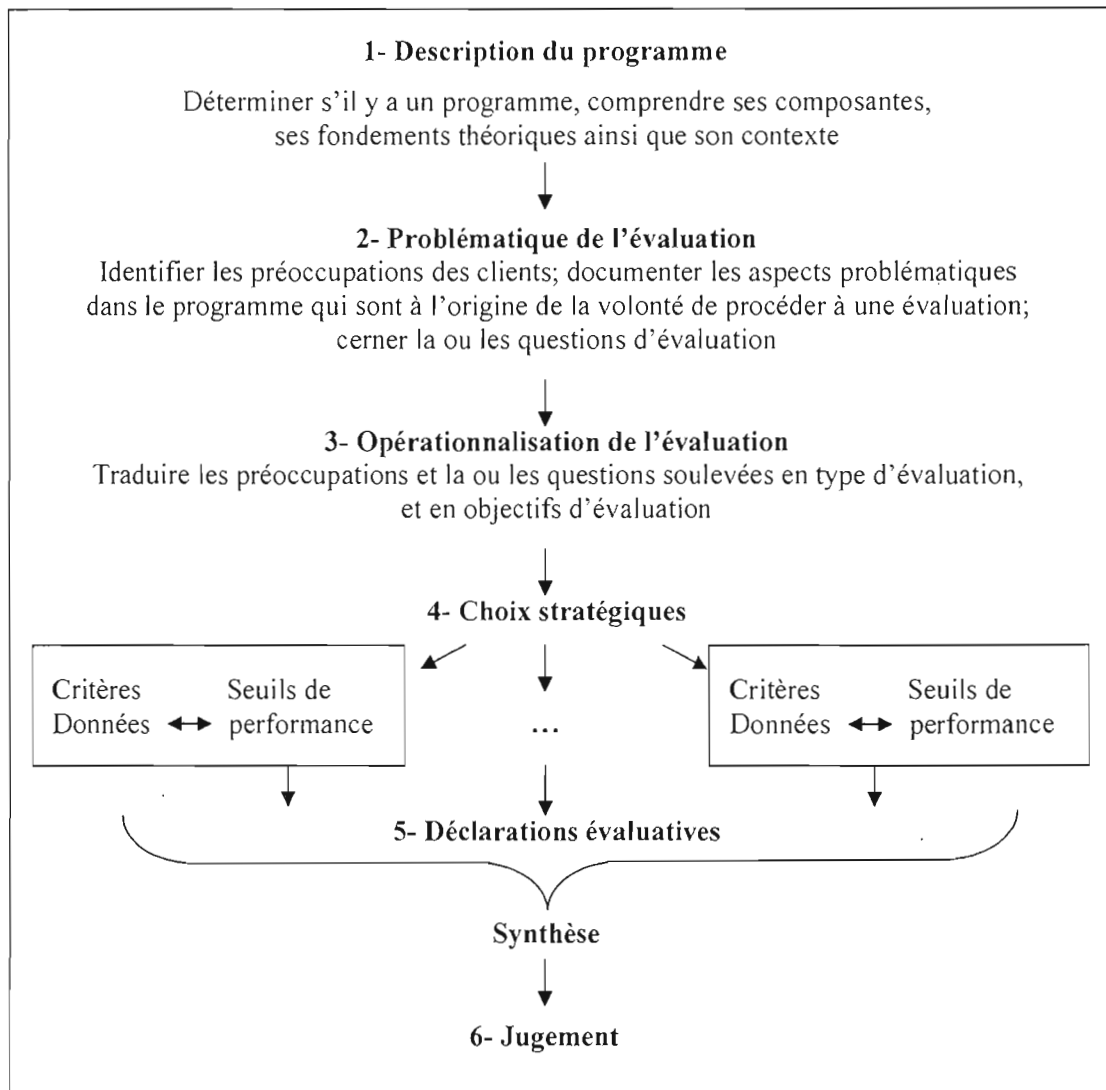


Figure 2.4 La modélisation du processus spécifique à l'évaluation de programme : composantes et interrelations. (Tirée de Hurteau, Lachapelle et Houle, 2006.)

Les six composantes de la modélisation sont présentées plus en détail ci-dessous :

- 1) La *description du programme* à évaluer consiste à amasser de l'information objective sur les caractéristiques du soi-disant programme afin de vérifier s'il en est réellement un.

- 2) La *problématique de l'évaluation* sert à identifier le facteur qui a déclenché l'évaluation (politique d'évaluation, problème rencontré, besoin d'information, etc.) ainsi que les préoccupations des clients de l'évaluation (Hurteau et Houle, 2006).
- 3) L'*opérationnalisation de l'évaluation* a pour fonction de structurer la démarche évaluative en tenant compte du contexte spécifique d'évaluation. Plus particulièrement, elle permet de traduire les besoins identifiés à l'intérieur de la problématique en question ou objectifs spécifiques d'évaluation, lesquels varient selon le type d'argumentation ou de jugement attendus (Fournier, 1995).
- 4) Les *choix stratégiques*, pierre angulaire de la démarche évaluative selon une majorité d'auteurs, consistent à définir les critères et les seuils de performance (éléments essentiels pour porter un jugement sur le programme).
- 5) Les *déclarations évaluatives* reposent sur l'analyse des écarts entre la mesure des performances observées et leurs seuils de performance respectifs afin de pouvoir discerner et mesurer efficacement les facteurs qui influencent les résultats obtenus (Hurteau, 1991; Fournier, 1995; Hurteau, Lachapelle et Houle, 2006; Arens, 2006), puis ultimement, de produire des déclarations justifiées.
- 6) Le *jugement* consiste à synthétiser les déclarations évaluatives afin d'émettre une conclusion évaluative, concise et cohérente, sur la valeur de l'objet d'évaluation du programme à l'étude (Scriven, 2006).

En 2006, cette modélisation a fait l'objet d'une validation (Hurteau, Lachapelle et Houle, 2006) et en 2007, son caractère générique a été établi dans le cadre de l'étude de Zielinski (2007). Il est à noter que même si l'illustration de ce processus apparaît linéaire, son application n'en demeure pas moins itérative.

En poursuivant leurs travaux de recherche, Hurteau et Houle (2005b) ont été mis devant l'évidence que si la modélisation expliquait les étapes requises pour générer le jugement, elle n'expliquait pas pour autant le processus sous-jacent (comment s'effectue le passage d'une étape à une autre). C'est ce qui a amené les chercheurs à s'intéresser au raisonnement évaluatif. Ainsi, la modélisation et le raisonnement sont deux processus complémentaires.

2.2.2.3 Le raisonnement évaluatif (Arens, 2006)

Pour parvenir à produire une conclusion évaluative valide, l'évaluateur suit un raisonnement propre à la pratique de l'évaluation de programme, à savoir le raisonnement évaluatif (Taylor, 1961; McCarthy, 1973, 1979; Toulmin, Rieke et Janik, 1984; Redding, 1989 et Toulmin, 1993). Au cours de ce raisonnement, l'évaluateur aura à produire un jugement (Valovirta, 2002), lequel se trouve au cœur de la démarche évaluative. Arens (2005 ; 2006), inspirée par les travaux de Toulmin, Rieke et Janik (1984), de Toulmin (1993), et de Fournier (1995), explique le développement du jugement au moyen d'un processus logique comprenant deux volets distincts, mais interreliés, soit la production d'un jugement et son argumentation. Cette démarche est illustrée au moyen de la figure 2.5.

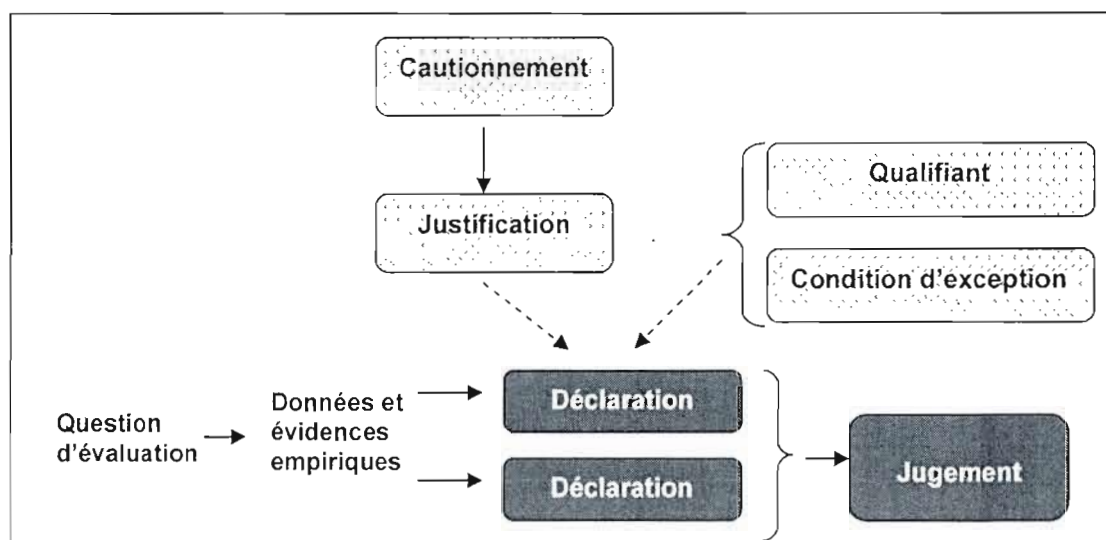


Figure 2.5 La démarche sous-jacente au raisonnement évaluatif : l'élaboration du jugement et de son argumentation. (Tirée d'Arens, 2006.)

Le premier volet du raisonnement évaluatif débute avec une question d'évaluation. Suite à la question posée, l'évaluateur recueille les informations pertinentes (données et évidences empiriques), puis il les synthétise en *déclarations*. Celles-ci sont des énoncés servant à la production d'un jugement (Valovirta, 2002). Le second volet du raisonnement évaluatif, soit l'argumentation, enrichit le premier. Afin que le jugement résiste à la critique, l'évaluateur

doit appuyer ses déclarations sur une argumentation solide (Toulmin, 1993) en tenant compte du contexte, des circonstances, des personnes impliquées, des ressources disponibles, etc. (House et Howe, 1999; Fournier, 1995; Arens, 2006). C'est à ce moment que le jugement devient crédible (Valovirta, 2002).

Plus particulièrement, la *justification* réfère aux prémisses (règles, faits, preuves, principes, critères, seuils de performance) qui servent de pont entre les déclarations, et les données et évidences produites (Toulmin, 1993). Elle permet, selon sa qualité, d'expliquer la légitimité et la validité de la déclaration, autrement dit, de comprendre comment l'évaluateur en arrive là (Toulmin, 1993). Derrière la justification se trouve normalement une autre assurance, soit le *cautionnement*, lequel supporte cette justification dans l'éventualité où elle serait remise en question. Il s'agit des fondements de la justification, c'est-à-dire des données additionnelles qui lui assurent autorité et crédit. À titre d'exemple, il peut s'agir d'un système taxonomique, d'un système de lois, de résultats d'études. Le *qualifiant* est un commentaire général qui pondère la portée de la déclaration en précisant les circonstances ou les conditions dans lesquelles la déclaration se vérifie, donc qui la rend crédible. La *condition d'exception* explicite la déclaration de façon plus contextuelle en fournissant les éventuelles conditions dans lesquelles elle pourrait être affaiblie ou ne pas être valide.

D'une certaine manière, l'évaluation de programme est comparable à l'enquête criminelle (Smith, 1995) qui consiste, elle aussi, à recueillir de l'information pertinente et rigoureuse (données et évidences) et à la synthétiser sous la forme d'une déclaration. Il y a vingt ans, Scriven (1981) avait déjà établi un rapprochement entre l'évaluation de programme et le plaidoyer d'un avocat devant des jurés. L'évaluateur doit convaincre les clients à accepter son jugement en apportant, au bon moment, des arguments rigoureux. Effectivement, un même argument peut être exposé sous un nombre considérable de formes différentes. Certains évaluateurs montreront plus clairement que d'autres la validité d'un argument et rendront plus explicites les justifications sur lesquelles s'appuient leurs déclarations (Toulmin, 1993). C'est lorsqu'un évaluateur possède cette compétence que les chances de rendre un jugement plus convaincant augmentent. Ainsi, la démarche évaluative

se trouve à être à la fois une science (collecte d'informations) et un art (l'habileté à bien argumenter) (Schwandt, 1990; Toulmin, 1993).

2.2.3 Conclusion sur l'impact des développements théoriques sur la pratique

Dans cette dernière section, nous avons présenté différents développements théoriques qui servent à encadrer et à supporter la pratique évaluative. Ainsi, en ce qui concerne *l'information requise pour appuyer le jugement*, Stake et Schwandt (2006) amènent les concepts de *qualité mesurée* et de *qualité appréciée*: Selon eux, les données quantitatives et qualitatives recueillies par le premier type de qualité favorisent la légitimité scientifique du jugement, tandis que les perceptions et opinions des détenteurs d'enjeux recueillies par le deuxième type de qualité contribuent à politiser le jugement. Bien que la proportion de chacune des qualités peut varier selon la nature et l'ampleur du programme à l'étude, l'évaluateur qui recourt aux deux types de qualité de façon complémentaire parviendrait à générer un jugement satisfaisant aux yeux des détenteurs d'enjeux. C'est ce qui fait dire à Brousselle, Champagne et Contandriopoulos (2006, p.68) que l'évaluation de programme devient donc « ...la rencontre de deux champs, l'espace politique et l'espace scientifique, entre lesquels il doit y avoir une contamination ».

Aussi, nous avons vu que les auteurs ne se sont pas seulement penchés sur l'information qui appuie le jugement, mais qu'ils ont aussi élaboré *différents processus qui encadrent la démarche évaluative* et aident l'évaluateur à produire un jugement crédible. À ce sujet, nous avons présenté la *modélisation de l'acte spécifique d'évaluer* d'Hurteau, Lachapelle et Houle (2006) ainsi que la *modélisation du raisonnement évaluatif* d'Arens (2006), deux processus complémentaires servant de cadre de référence pour guider l'évaluateur. Ceux-ci consistent à donner de la légitimité à la démarche évaluative.

2.3 Synthèse et objectifs de recherche

Il existe une multitude de chemins pour se rendre de la définition du problème au jugement évaluatif. La difficulté majeure consiste à trouver la meilleure voie à prendre pour rendre le jugement crédible (Nevo, 2006). Ainsi, au regard des éléments approfondis à l'intérieur de cette section, nous avons vu que, depuis les années 1970, de nombreux modèles, méthodes, techniques et théories propres à l'évaluation de programme ont été développés pour aider les évaluateurs. À titre d'exemple, en ce qui a trait au processus, la modélisation de l'évaluation de programme de Hurteau et Houle (2005b) ainsi que le raisonnement évaluatif d'Arens (2006) offrent, à eux deux, une structure de travail pour faciliter la formulation d'un jugement crédible. Toutefois, tel que mentionné précédemment, ces processus s'avèrent souvent inopérants dans la pratique, lorsqu'ils ne se retrouvent pas simplement ignorés des évaluateurs. Effectivement, comme le démontrent les deux constats de l'étude d'Arens (2006), les praticiens ne recourent pas de façon efficace aux structures du raisonnement évaluatif pour guider leur pratique. Dans le même ordre d'idées, Stake et *al.* (1997) remarquent, eux aussi, que la logique de l'évaluation et son processus d'analyse et de synthèse ne trouvent pas d'écho dans la pratique. Suite à leur étude, Hurteau, Houle et Mongiat (à paraître) arrivent à la conclusion que les évaluateurs ne semblent pas suivre les étapes spécifiques de leur modélisation. Finalement, l'étude de Christie (2003) révèle que les modèles théoriques sont rarement utilisés comme il se doit dans la pratique.

Toutes ces études démontrent que les évaluateurs ne se réfèrent pas aux théories formelles développées en évaluation, du moins lorsque leurs rapports d'évaluation sont analysés. Les théoriciens ne peuvent donc pas encore établir de liens directs entre les éléments ou les processus qui mènent à la production d'un jugement crédible. Bref, l'ensemble des développements théoriques portant sur le développement de l'argumentation générant le jugement évaluatif semble mener à une impasse. Ce constat amène Christie (2003, p.92) à se questionner : « [...] if formal theory does not guide their work, then what does? Research on evaluation can help answer this question. »

De la même façon, l'élaboration de notre problématique de recherche nous a conduits à la formulation de la question suivante : *De façon générale, quels sont les éléments et les procédures liés à la pratique évaluative qui sont requis pour générer un jugement satisfaisant aux yeux des clients de l'évaluation?*

Le cadre théorique de ce présent mémoire a présenté de nouvelles perspectives qui portent un regard différent sur la façon de générer un jugement satisfaisant aux yeux des clients de l'évaluation. Ainsi, plusieurs auteurs parlent de l'importance d'identifier la façon dont les résultats de l'évaluation seront récupérés dans la pratique et d'un jugement qui tient compte non seulement des faits, mais aussi des valeurs (Shadish, Cook et Leviton, 1991; Alkin et Chrstie, 2004). D'autres ont une vision de l'évaluation plus démocratique dans laquelle l'évaluateur a un rôle de négociateur et dans laquelle il est important de prendre en considération les différentes perspectives des détenteurs d'enjeux quand vient le temps d'élaborer le jugement (Stufflebeam, 2000; House et Howe, 1999). Ainsi, l'évaluateur recueille de nouvelles informations qui lui permettent de développer une meilleure compréhension du programme, et ultimement, de rendre son jugement plus acceptable. Quant à Stake et Schwandt (2006), ils amènent les concepts de qualité mesurée et de qualité appréciée. À leur avis, l'évaluateur qui recourt aux deux types de qualité de façon complémentaire pour élaborer son jugement parvient à donner un maximum de crédibilité à ce dernier. À ce propos, Stake et Schwandt (2006) soulignent qu'une des façons de recueillir de l'information basée sur la qualité appréciée est d'impliquer les détenteurs d'enjeux à l'ensemble de la démarche évaluative.

Toutefois, la conception actuelle de l'évaluation de programme est encore très limitée en ce qui concerne la reconnaissance accordée à la qualité appréciée dans la démarche évaluative (House et Howe, 1999). Par le fait même, très peu d'auteurs traitent de ce concept. D'ailleurs, les différents *processus qui encadrent la démarche évaluative* représentent une limite en ce qui a trait à l'argumentation qui sous-tend le jugement puisqu'ils ne prennent pas en considération la *qualité appréciée*. Ils contribuent donc uniquement au volet de la légitimité scientifique. Ainsi, les modélisations discutées dans le cadre théorique s'avèrent inefficaces pour intégrer les deux types d'information afin de produire un jugement jugé

satisfaisant aux yeux des clients. Est-ce que cette lacune contribue aux remises en question auxquelles l'évaluation fait face ?

Bien que les théoriciens ne s'intéressent pas encore à la qualité appréciée, qu'en est-il des praticiens? Il est impératif d'aller voir ce que les principaux concernés en pensent. Toutes ces considérations nous permettent de cibler quatre objectifs, lesquels rendront notre question de recherche opérationnelle :

- 1) Établir ce qu'est un jugement satisfaisant;*
- 2) Identifier les éléments contribuant à générer un jugement satisfaisant;*
- 3) Explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant;*
- 4) Explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant;*

et ce, selon les perspectives d'évaluateurs et de clients de l'évaluation.

À travers le cadre théorique, nous avons développé les différents concepts en jeu, soit l'évaluation de programme ainsi que les composantes essentielles à la production d'un jugement crédible. Cette démarche nous a non seulement permis de mieux cerner la problématique de notre projet de recherche, mais elle nous a également guidé vers des choix méthodologiques qui nous aideront à atteindre nos objectifs de recherche. Les détails en ce qui concerne notre démarche méthodologique seront abordés au prochain chapitre.

CHAPITRE III

DÉMARCHE MÉTHODOLOGIQUE

Ce troisième chapitre décrit le détail de la méthodologie qui vise la vérification des objectifs de recherche énoncés dans le cadre théorique. Tout d'abord, le type de recherche et la justification des choix méthodologiques seront présentés, incluant la description de l'échantillon des sujets qui participeront à la recherche et la méthode de collecte de données retenue. Ensuite, la présentation de la méthode d'analyse et d'interprétation des données sera abordée sommairement ainsi que les modalités pour assurer une rigueur à la démarche. Finalement, cette section se terminera avec la présentation des forces et des limites de l'étude.

3.1 Le type d'informations requises

La présente recherche a pour objet d'étude les éléments et les procédures qui génèrent un jugement évaluatif satisfaisant aux yeux des clients de l'évaluation. Le cadre théorique de notre étude nous a révélé que l'évaluateur ne devrait pas seulement faire appel à la qualité mesurée pour argumenter son jugement, mais qu'il devrait recourir également à la qualité appréciée. Nous ne connaissons pas la perception qu'ont les praticiens face à ces deux types de qualité. Cette information n'est d'ailleurs pas toujours visible à l'intérieur des rapports d'évaluation. Il ne serait donc pas pertinent de procéder à une analyse de contenu comme cela a été fait dans les deux premiers volets du programme de recherche d'Hurteau et Houle (2005). Étant donné que notre recherche se centre davantage sur le raisonnement, il nous faut plutôt recueillir de l'information directement auprès des personnes concernées par l'évaluation, soit les évaluateurs et les clients.

3.2 Le type de recherche

Tout d'abord, commençons par rappeler les quatre objectifs qui ont été retenus suite à l'élaboration de notre cadre théorique :

- 1) Établir ce qu'est un jugement satisfaisant;*
- 2) Identifier les éléments contribuant à générer un jugement satisfaisant;*
- 3) Explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant;*
- 4) Explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant;*

et ce, selon les perspectives d'évaluateurs et de clients de l'évaluation.

Considérant la nature de ces objectifs, la présente étude s'inscrit dans la perspective d'une recherche fondamentale de type exploratoire se situant dans un contexte de découverte (Lessard-Hébert, Goyette et Boutin, 1990).

3.3 Les choix méthodologiques

Les choix méthodologiques seront maintenant présentés sous deux aspects, soit le choix de l'échantillon et la méthode de collecte de données.

3.3.1 L'échantillon

Les participants ont été sélectionnés par la technique d'échantillonnage intentionnel parce que nous voulions joindre une clientèle précise, soit des évaluateurs chevronnés et des clients de l'évaluation possédant les caractéristiques suivantes : 1) familiers avec le processus de

l'évaluation; 2) impliqués dans le domaine; 3) ayant des connaissances et une expertise dans le domaine, surtout en ce qui concerne les évaluateurs. En résumé, les répondants devaient être associés à des démarches évaluatives sur une base régulière depuis plusieurs années, que ce soit à titre d'évaluateur ou de client. En choisissant nous-mêmes les participants, nous nous assurons qu'ils correspondaient exactement à nos critères de sélection et, par le fait même, nous maximisions les chances de construire un corpus de données riches (Karsenti et Savoie-Zajc, 2004). De plus, le fait de sélectionner deux catégories de répondants nous amenait à rassembler différentes perspectives sur ce qui fait qu'un jugement est satisfaisant. Ce choix nous permettait d'éventuellement nuancer nos résultats de recherche.

Au total, six personnes ont été recrutées pour cette recherche, soit trois évaluateurs chevronnés et trois clients de l'évaluation qui possédaient tous les critères recherchés. Ces participants œuvraient dans différents types d'organisations de la région de Montréal et d'Ottawa : organismes gouvernementaux et paragouvernementaux, organismes communautaires et universitaires.

3.3.2 La méthode de collecte de données

Considérant l'objet d'étude, soit les éléments et les procédures qui assurent un jugement évaluatif satisfaisant, comme un phénomène complexe qui n'est pas entièrement et directement observable, surtout lorsque nous parlons de perspectives, les entrevues en personne constituaient une méthode pertinente pour recueillir nos données (Karsenti et Savoie-Zajc, 2004). En effet, l'entrevue en personne s'avère la méthode la plus appropriée lorsqu'il s'agit de recueillir des données sur l'expérience des participants et que cela constitue des impressions, des croyances et des opinions (Lessard-Hébert, Goyette et Boutin, 1990). Ainsi, grâce à cette méthode de collecte de données, il nous a été possible de laisser émerger et d'approfondir les points de vue qu'entretiennent les participants face au jugement. En ayant un contact direct avec les répondants, nous pouvions ajuster ou reformuler nos questions pour qu'ils en comprennent bien le sens (Boutin, 1997). De plus, Boutin (2006)

ajoute que la plupart du temps, l'entrevue est la meilleure méthode pour recueillir des informations riches.

Nous avons opté pour des entrevues individuelles afin de permettre aux participants d'articuler librement leur point de vue sur des aspects qui devaient émerger de leur pratique. Dans ce sens, l'entrevue de groupe n'a pas été considérée puisque nous voulions éviter que leur point de vue soit contaminé par d'autres perspectives. Comme le mentionne Boutin (1997), les entrevues individuelles réduisent les chances que les participants s'influencent mutuellement ou encore, qu'ils donnent les réponses auxquelles le chercheur s'attend pour bien paraître devant le reste du groupe. De plus, la peur d'être jugé ou la réticence à dévoiler certaines informations personnelles devant des collègues n'était pas à négliger.

L'entrevue semi-structurée nous est apparue pertinente à utiliser dans la présente étude puisqu'elle permettait aux participants de développer leur point de vue sur des questions larges. De plus, la souplesse dans la formulation et la séquence des questions nous donnaient la possibilité de nous réajuster en temps réel en demandant aux participants, par exemple, d'élaborer sur une question en particulier.

À ce propos, nous retrouvons à l'appendice A le questionnaire d'entrevue que nous avons élaboré à partir des thèmes et des sous-thèmes centraux émanant du cadre théorique (Karsenti et Savoie-Zajc, 2004). Des questions ouvertes nous ont permis de guider les échanges des entrevues afin de faire émerger le vécu personnel des répondants, de stimuler leur réflexion et de couvrir les thèmes à explorer.

Afin de valider le contenu (pertinence des questions) et la forme de notre questionnaire (clarté et ordre des questions), nous avons préalablement soumis une première version du questionnaire à deux évaluateurs experts, également professeurs à l'Université du Québec à Montréal. Les commentaires émis par ces professeurs ont mené à l'élaboration d'une deuxième version du questionnaire, laquelle a été présentée aux membres du comité d'évaluation de l'avant-projet de ce mémoire. Par la suite, le questionnaire a été administré à une tierce personne afin de s'assurer de la clarté des questions (revêtent-elles le sens qu'elles

devraient?). Comme le questionnaire et son processus d'administration ne semblaient pas poser problème, nous avons alors procédé aux réalisations des entrevues.

En résumé, nous avons procédé à un échantillonnage intentionnel pour recruter nos participants. Au total, notre échantillon est composé de trois évaluateurs chevronnés et de trois clients de l'évaluation. Nous avons procédé à des entrevues individuelles, semi-structurées pour recueillir nos données.

3.4 L'analyse et l'interprétation des données

Les entrevues ont été analysées à partir des objectifs de recherche établis par les éléments théoriques documentés dans le cadre théorique. Cette démarche s'inscrivait dans une perspective d'*analyse inductive générale*, procédure de réduction de données proposée par Thomas (2006). Ce type d'analyse d'entrevues convient particulièrement aux traitements de données qualitatives en recherche exploratoire.

Blais et Martineau (2006) décrivent l'analyse inductive comme une procédure qui consiste en plusieurs *lectures attentives et approfondies* des données brutes (*verbatim*s des entrevues présentés dans un format uniforme et analysable) dans le but d'identifier les segments de textes significatifs (unités de sens), puis de les regrouper dans des catégories généralement liées aux objectifs ou aux questions de la recherche. Par la suite, il est souvent nécessaire d'effectuer un *raffinement des catégories*. Pour ce faire, le chercheur doit fusionner ou modifier les catégories préalablement identifiées pour ne disposer, au total, que d'un maximum de huit catégories.

C'est le logiciel QDA Miner qui a été utilisé pour traiter les *verbatim*s étant donné qu'il se prête bien au traitement d'entrevues.

3.5 Les considérations éthiques

Les sujets ont participé de façon volontaire à l'étude et ils étaient libres de se retirer de l'entrevue à tout moment s'ils le désiraient. Les participants ont clairement été informés à l'avance du sujet de la recherche, de ses objectifs et de son déroulement. En outre, chacun d'entre eux a dû signer un formulaire d'information et de consentement (voir appendice B). Tel que souhaité par les participants et selon les règlements institutionnels en vigueur, l'anonymat et la confidentialité des informations recueillies ont été respectés lors du traitement et de l'analyse des données ainsi que lors de la diffusion des résultats de recherche. À cet effet, la présente étude a fait la demande d'un certificat de déontologie. De plus, un code a servi à identifier les sujets de telle sorte qu'aucune information permettant de les identifier n'apparaissait dans les bases de données. Les renseignements ont été conservés dans un classeur sous clé situé dans un bureau fermé. Tous les enregistrements des entrevues ont été détruits à la fin de la recherche afin de garantir la confidentialité de l'information. Finalement, une fois la présente étude terminée, les participants ont été informés des résultats de recherche via leur adresse de courrier électronique.

3.6 Les forces et les limites de la recherche

Différents choix méthodologiques effectués dans cette recherche ont permis de nous assurer de l'atteinte de nos objectifs. Nous présentons ci-dessous, les deux principales forces méthodologiques de notre étude. Puis, nous terminerons ce chapitre en présentant les limites méthodologiques de cette étude.

- Les forces de la recherche

Premièrement, la présente étude revêt un certain intérêt par son effort à dépasser les spéculations traditionnelles concernant les pratiques évaluatives. Au moyen d'entrevues, nous

sommes allés chercher directement les perspectives d'évaluateurs chevronnés et de clients de l'évaluation concernant le jugement qui les satisfait.

Deuxièmement, nous avons procédé à une transcription intégrale des six entrevues individuelles, puis nous avons analysé ces *verbatim*s à l'aide d'un logiciel d'analyse de données qualitatives reconnues par le milieu scientifique, soit QDA Miner. Ce processus apporte de la rigueur à notre démarche méthodologique, et par conséquent à nos résultats de recherche.

- Les limites de la recherche

Tout d'abord, malgré le fait que notre étude s'appuie sur un cadre théorique, celui-ci demeure peu étoffé parce que l'évaluation de programme ne fait pas l'objet de beaucoup de recherches. Ensuite, il n'a pas été possible de soumettre le questionnaire d'entrevue à un troisième évaluateur expert pour discuter de son contenu et départager les positions, car peu de personnes possèdent l'expertise nécessaire pour réagir à un contenu aussi précis.

CHAPITRE IV

RÉSULTATS

Ce quatrième chapitre est consacré aux résultats de la recherche. Rappelons que trois évaluateurs chevronnés et trois clients ayant vécu le processus d'une évaluation ont été sélectionnés de façon intentionnelle pour prendre part à nos entrevues individuelles. Ce chapitre est constitué de deux parties distinctes : la première reprend les différentes étapes auxquelles nous avons eu recours lors de l'analyse des données, tandis que la deuxième partie consiste à présenter les résultats de recherche en tant que tels, lesquels permettront de répondre à nos objectifs de recherche.

4.1 L'analyse des données

Avant de présenter nos résultats de recherche, il est important de préciser la démarche que nous avons suivie pour analyser les données recueillies lors des entrevues individuelles. Deux grandes étapes composent cette analyse, soit le codage des données et la synthèse des données.

4.1.1 Le codage des données

Une fois les six entrevues réalisées et transcrites en *verbatim*s, nous avons procédé à une première lecture du matériel (données brutes) pour avoir une connaissance globale des sujets couverts. Par la suite, à l'aide du logiciel QDA Miner, nous avons lu attentivement une deuxième fois les *verbatim*s afin de relever les extraits de texte qui suscitaient un intérêt particulier (unités de sens). L'étape de la détermination des unités de sens revêt une dimension subjective, car il nous faut faire un choix quant au matériel retenu pour le codage et le traitement des données. Par contre, cette subjectivité fait partie du processus même de l'analyse. De plus, comme le souligne Boutin (2006), le jugement posé sur l'importance de

tel ou tel passage dépend de l'expérience du chercheur, de sa sensibilité aussi bien que de la connaissance qu'il a du sujet.

À partir des lectures répétées du contenu de chacune des entrevues, il a été possible de dégager des catégories et d'ensuite y regrouper les unités de sens précédemment identifiées (codage des données). Notons que certaines catégories ont été élaborées à partir du cadre théorique de ce mémoire, mais que d'autres ont émergé lors du codage des *verbatim*s.

Le logiciel QDA Miner nous a permis d'attribuer une couleur pour chacune des catégories facilitant ainsi le repérage des unités de sens puisque celles-ci ont été codées de la même couleur que la catégorie dans lesquelles elles s'inséraient. Il est à noter que ce logiciel nous a aussi été utile pour repérer, classer et organiser les données. Une fois le codage des *verbatim*s terminé, nous avons pu procéder à la synthèse des données.

4.1.2 La synthèse des données

La synthèse des données, deuxième grande étape de l'analyse des données, se subdivise en deux sous-étapes, soit l'examen des données et le traitement des données.

4.1.2.1 L'examen des données

L'examen des données consiste à examiner les unités de sens relevées à l'étape précédente afin d'obtenir un portrait global des données. À l'aide du codage inverse (fonction dans QDA Miner), une lecture des unités de sens nous a permis de nous assurer que toutes ces unités étaient correctement liées aux différentes catégories dans lesquelles elles se retrouvaient. De plus, c'est également à ce moment que nous avons examiné si les unités de sens sélectionnées nous fournissaient l'information pour répondre à nos objectifs de recherche de façon adéquate.

Le recours à des tableaux de fréquences (0-1) fournis par le logiciel QDA Miner nous a permis d'illustrer de façon visuelle l'ensemble de ces données. Le choix de ce procédé permet au chercheur d'examiner et de comparer en un coup d'œil la distribution des réponses (codes) en fonction du nombre de répondants (Huberman et Miles, 2003). Ainsi, si une catégorie est plus ou moins utilisée, il peut être intéressant d'y porter une attention particulière. Toutefois, nous sommes conscients que notre recherche se situe dans une perspective exploratoire. Notre échantillon se limitant à six participants, il est difficile, dans ce contexte-ci, d'accorder une trop grande importance aux fréquences. C'est la raison pour laquelle nous avons choisi de ne pas présenter ces tableaux de fréquence dans ce chapitre.

4.1.2.2 Le traitement des données

L'étape suivante est consacrée à la réduction des données dans le but de pouvoir présenter les résultats sous une forme plus accessible. Cette étape, où pour chacune des catégories, on structure et condense les extraits de *verbatim* (unités de sens) les plus significatifs répertoriés à l'étape précédente, a pour but de pouvoir procéder à une analyse des données plus approfondie. Cette opération permet d'établir des liens entre les contenus des différentes entrevues effectuées auprès des participants. C'est ainsi que nous nous sommes rendu compte que certaines catégories avaient des unités de sens trop similaires. Nous avons alors procédé à des regroupements de catégories plutôt qu'à l'élimination d'une des catégories similaires.

Toutes ces étapes ont ainsi constitué une phase préparatoire indispensable à la présentation des résultats, section traitée à la page suivante. Au besoin, des citations (extraits du *verbatim*) seront utilisées pour appuyer nos résultats, et ce, dans le but de faire ressortir davantage la teneur des données recueillies ou pour ne pas perdre le sens véritable des différents propos émis par les sujets concernés.

4.2 La présentation des résultats de recherche

Ici sont présentés les résultats issus de l'analyse des *verbatim* des six entrevues individuelles que nous avons faites. Afin de ne pas influencer les personnes interrogées, notre questionnaire d'entrevue comportait certaines questions génériques portant sur le rapport d'évaluation. Toutefois, dans le présent chapitre, nous ne traiterons que de ce qui se rapporte au jugement en tant que tel puisque notre question de recherche va dans ce sens : *De façon générale, quels sont les éléments et les procédures liés à la pratique évaluative qui sont requis pour générer un jugement satisfaisant aux yeux des clients de l'évaluation?*

Rappelons également que cette question a entraîné la formulation de quatre objectifs de recherche. D'ailleurs, la présentation des résultats est organisée selon chacun de ces objectifs :

- 1) *Établir ce qu'est un jugement satisfaisant;*
 - 2) *Identifier les éléments contribuant à générer un jugement satisfaisant;*
 - 3) *Explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant;*
 - 4) *Explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant;*
- et ce, selon les perspectives d'évaluateurs et de clients de l'évaluation.*

4.2.1 Objectif 1 : établir ce qu'est un jugement satisfaisant

Il se dégage de l'analyse des données trois critères pour établir ce que constitue un jugement satisfaisant aux yeux des clients de l'évaluation. Ainsi, selon les évaluateurs et les clients interviewés, le jugement doit être argumenté, utile et explicite.

4.2.1.1 Description des critères caractérisant un jugement satisfaisant

Nous allons donner une courte définition pour chacun des trois critères et nous pencher sur les raisons invoquées par les participants quant à l'importance d'un jugement possédant ces qualités. Notez que tous les résultats ci-dessous ont été formulés à partir d'une synthèse des réponses données par les répondants lors des entrevues (voir section 4.1.2.2 sur le traitement des données).

- Un jugement argumenté

Pour tous les répondants (n=6), un jugement satisfaisant doit être avant tout bien argumenté, c'est-à-dire qu'il doit s'appuyer sur de l'information de qualité qui est présentée adéquatement. Ce constat réfère donc à deux aspects, d'une part le *contenu* et, d'autre part, la *présentation*. En ce qui concerne le *contenu*, bien que les répondants valorisent des informations abondantes et décrites en profondeur, ils veulent avant tout qu'elles réfèrent systématiquement à l'objet évalué. Autrement dit, les informations, quel qu'en soit le volume, doivent être pertinentes et éclairer le lecteur sur la teneur du jugement. Bien que les répondants ne l'aient pas mentionné dans des termes aussi clairs, ils précisent que pour recueillir toute l'information pertinente nécessaire pour argumenter le jugement, l'évaluateur devrait recourir de façon équilibrée à la qualité mesurée et à la qualité appréciée. Nous tenons à faire remarquer ici que l'importance de nuancer le jugement est ressortie dans le discours de tous les participants. Selon eux, si un jugement doit nécessairement s'appuyer sur des éléments relevant de la qualité mesurée, ne serait-ce que pour lui fournir une légitimité scientifique, il ne peut faire abstraction de la qualité appréciée.

En ce qui concerne la *présentation*, l'évaluateur doit porter une attention particulière à la façon dont il amène ses arguments afin de bien les mettre en valeur et de leur donner un maximum de poids. Savoir agencer les arguments provenant de la qualité mesurée et de la

qualité appréciée pour rendre le jugement le plus crédible et convaincant possible est un art en soi.

En somme, pour les répondants, il est important que le jugement soit argumenté afin:

- de s'éloigner des jugements de valeur;
- d'obtenir un rapport qui reflète une réalité plus juste;
- d'aider les clients à mieux comprendre le jugement et de retracer ses fondements (à quoi il s'accroche);
- d'émettre un jugement logique et défendable;
- d'augmenter la crédibilité du jugement;
- d'assurer un jugement, acceptable et accepté de tous, bref de convaincre le client du bien-fondé du jugement.

- Un jugement utile

Toujours selon les évaluateurs et les clients interviewés, un jugement satisfaisant doit aussi être utile aux détenteurs d'enjeux. Un seul répondant, un évaluateur plus précisément, n'a pas abordé spontanément cet aspect au cours des entrevues (n=5). Un client a souligné l'importance de sentir à travers le jugement émis que « l'évaluateur n'a pas seulement remis un rapport pour remettre un rapport », mais qu'il a tenu compte des éventuelles répercussions (positives ou négatives) que son jugement pouvait avoir sur les suites du programme. Ainsi, l'évaluateur doit faire preuve de pédagogie et amener son jugement de façon constructive. De plus, il faut qu'il s'efforce de produire un jugement qui correspond aux besoins des détenteurs d'enjeux et qui répond à leurs attentes.

En somme, il est important que le jugement soit utile afin:

- de maximiser les chances d'utilisation du rapport;
- d'éclairer la prise de décision quant aux orientations futures du programme;
- d'augmenter les connaissances sur le programme en question;
- de pouvoir tirer des leçons.

- Un jugement explicite

Finalement, quatre répondants sur six, deux évaluateurs et deux clients, trouvent essentiel que le jugement soit explicite. Ceux-ci parlent d'un jugement qui est clair et catégorique dans lequel ils ne retrouvent aucune ambiguïté. Ainsi, l'évaluateur ne doit laisser aucune question en suspens quant à la teneur du jugement. De plus, il doit bien identifier la section « jugement » ou « conclusion » dans son rapport afin de la distinguer facilement des autres sections, y compris de celle des « recommandations ».

En somme, il est important que le jugement soit explicite afin:

-d'éviter que le client n'ait besoin de lire entre les lignes ou de chercher le jugement.

4.2.1.2 Conclusion de l'objectif 1 : établir ce qu'est un jugement satisfaisant

En conclusion, nous constatons que de façon générale, les évaluateurs et les clients interviewés s'entendent pour dire qu'un jugement satisfaisant doit regrouper trois caractéristiques pour être considéré comme tel. Le jugement doit être : 1) argumenté, c'est-à-dire qu'il appuie ce qui est avancé par des arguments solides et nuancés; 2) utile, dans le sens où il sert aux détenteurs d'enjeux pour prendre des décisions quant aux orientations futures du programme; 3) explicite, dans le but qu'il soit facilement compréhensible par les personnes qui reçoivent le rapport.

Une fois que les critères qui caractérisent un jugement satisfaisant ont été établis, il est possible de se questionner sur la façon d'arriver à produire un tel jugement. C'est le deuxième objectif de notre étude qui tentera d'apporter des éléments de réponse à cette question.

4.2.2 Objectif 2 : identifier les éléments contribuant à générer un jugement satisfaisant

Selon les évaluateurs et les clients interviewés, il se dégage de l'analyse des données cinq éléments qui contribuent à générer un jugement satisfaisant aux yeux des clients de l'évaluation, soit : la question et/ou les objectifs d'évaluation, la description du programme, la méthodologie, l'information recueillies ainsi que la cohérence. Tandis que les quatre premiers éléments réfèrent au contenu du rapport, le dernier, soit la cohérence, se rapporte à sa forme. Tout d'abord, nous décrirons ces éléments, puis nous présenterons les éléments pour lesquels les répondants ont accordé la plus grande pondération.

4.2.2.1 Description des éléments contribuant à générer un jugement satisfaisant

Nous allons donner une courte définition pour chacun des cinq éléments qui contribuent à générer un jugement satisfaisant ainsi qu'expliquer comment ils se construisent. Rappelons que tous les résultats ci-dessous ont été formulés à partir d'une synthèse des réponses données par les répondants lors des entrevues (voir section 4.1.2.2 sur le traitement des données).

- La question d'évaluation et/ou les objectifs d'évaluation

Tous les répondants (n=6) ont mentionné que la question d'évaluation et les objectifs d'évaluation sont des éléments qui contribuent à générer un jugement satisfaisant. Étant donné que les descriptions qu'ils en faisaient se ressemblaient beaucoup, nous avons décidé de les regrouper à l'intérieur d'une même section plutôt que de les traiter séparément.

Dans un premier temps, l'évaluateur doit circonscrire et formuler clairement une question d'évaluation et/ou des objectifs d'évaluation opérationnels qu'il soumet aux clients de l'évaluation. Tous ensemble, ils discutent et négocient jusqu'à ce qu'il y ait consensus, et ce,

tant sur le plan de la problématique que sur le plan du choix des objectifs. Dans un deuxième temps, l'évaluateur doit définir clairement chacun des objectifs dans le rapport. Ainsi, il s'assure que les lecteurs aient une compréhension commune du mandat de l'évaluation.

C'est en impliquant les détenteurs d'enjeux que l'évaluateur s'assure de la pertinence de la question d'évaluation. Puisqu'ils connaissent mieux que quiconque la problématique sur laquelle il faut se pencher, il est important de les écouter. Quant aux objectifs, eux aussi risquent davantage de répondre aux attentes et aux besoins réels des clients s'ils sont élaborés en travaillant en collaboration avec eux. Par conséquent, l'évaluation prend un maximum de sens aux yeux des clients, et le jugement qui en découle s'en ressent aussi.

- La description du programme

Bien que le jugement se construise à partir de ce qui a été observé et analysé, deux évaluateurs sur trois précisent que le jugement doit aller au-delà de la méthodologie et de la procédure, c'est-à-dire que le jugement doit être ajusté au contexte du programme et être en relation directe avec celui-ci. Pour arriver à démontrer qu'il comprend bien le contexte du programme, l'évaluateur doit faire une description du programme à évaluer, dans laquelle il tiendra compte :

- des composantes du programme
- de la population visée par le programme
- de la façon dont l'implantation du programme a été faite et du déroulement des activités et services
- des aspects spécifiques de l'environnement du programme
- des enjeux du programme

En décrivant le programme de la sorte, l'évaluateur s'assure de démontrer aux clients qu'il a une vision juste du programme. Il pourra également s'en servir comme arguments pour pondérer et nuancer son jugement final.

- La méthodologie

Toutes les personnes interviewées (n=6) sont d'avis que la méthodologie a un impact direct sur la crédibilité du jugement et qu'elle assure la légitimité scientifique. Selon eux, il est important que l'évaluateur énonce clairement à l'intérieur du rapport les *choix méthodologiques* qui ont été faits (ex. : échantillon, outils de collecte) et qu'il explique la façon dont les données ont été recueillies (ex. : contexte de collecte). Toutefois, un des clients interviewés a été plus large dans sa définition de la méthodologie et a inclus les *choix stratégiques* à cet élément, tel que le démontre la citation suivante: « Aussi, dans la partie méthodologie du rapport, je veux retrouver les critères, les seuils, sur quoi porte le jugement... ». Par la suite, un autre client a insisté sur l'importance de s'assurer que les différentes perspectives des détenteurs d'enjeux soient prises en compte lors de cette étape. Puis, cinq répondants sur six ont parlé d'énoncer clairement les forces et limites de la démarche méthodologique. Celles-ci apportent un autre éclairage sur les résultats de l'évaluation, permettent parfois d'expliquer certaines zones grises du rapport, mais apportent surtout des nuances au jugement. Finalement, tous les répondants s'entendent pour dire que l'évaluateur doit faire preuve de rigueur dans l'application de sa méthodologie pour être en mesure de recueillir des informations qui reflètent le programme d'une façon juste.

- Les informations recueillies

Comme nous l'avons déjà vu, la qualité des informations recueillies est essentielle pour parvenir à générer un jugement bien argumenté qui saura satisfaire les clients de l'évaluation. Ainsi, tous les participants (n=6) sont d'avis que pour appuyer le jugement, l'évaluateur doit autant recourir aux données issues de sa collecte de données (qualité mesurée) que se servir des informations pertinentes qui ont émergé au cours de l'évaluation (qualité appréciée). Tous reconnaissent non seulement l'importance de recueillir les différentes perceptions, et ce, qu'elles proviennent des détenteurs d'enjeux ou de l'évaluateur lui-même, mais aussi d'en tenir compte quand vient le temps de porter un jugement sur le programme. Pour ce faire,

l'évaluateur doit consulter les personnes concernées par le programme et travailler en collaboration avec elles à toutes les étapes de l'évaluation. Suite à la multitude d'informations générées par la démarche évaluative, les répondants notent qu'il est souvent nécessaire que l'évaluateur procède à une sélection et à un triage afin d'en dégager les réels enjeux. Par la suite, son défi est d'intégrer et d'harmoniser les informations de différentes natures en vue de générer un jugement satisfaisant aux yeux des clients.

- La cohérence

Finalemt, pour presque tous les répondants (n=5), la cohérence est un élément essentiel qui contribue à générer un jugement satisfaisant. Un seul évaluateur n'a pas spontanément abordé cet aspect au cours des entrevues. Contrairement aux quatre autres éléments nommés précédemment, la cohérence ne fait plus référence au contenu du rapport, mais bien à sa forme.

Ainsi, l'évaluateur doit s'assurer que les différentes sections du rapport d'évaluation soient cohérentes entre elles (cohérence « verticale »), c'est-à-dire qu'il faut retrouver un fil conducteur tout au long de la lecture du rapport. Toutes ces différentes sections, bien que différentes, doivent se tenir entre elles, et ce, peu importe la taille du programme à évaluer.

Plus précisément, l'évaluateur doit vérifier si :

- la question d'évaluation et/ou les objectifs sont liés à la méthodologie;
- la méthodologie est appropriée au contexte du programme;
- les méthodes de traitement de données utilisées sont liées logiquement aux données recueillies;
- le jugement est fondé et soutenu par la méthodologie;
- le jugement émis répond à toutes les questions et/ou objectifs d'évaluation;

De plus, l'évaluateur doit aussi s'assurer que le contenu de chacune des sections soit cohérent (cohérence « horizontale »), c'est-à-dire que tous les éléments soient présents et se

raccrochent au contexte du programme. L'évaluateur doit éviter d'émettre des contradictions ou de l'information non pertinente dans son rapport afin d'éviter de *créer du bruit*.

Ainsi, un client qui lit un rapport cohérent peut retracer plus facilement les fondements du jugement parce que tout est bien structuré, fluide et logique. Un des répondants parle de l'importance de « prendre le lecteur par la main afin d'éviter qu'il ne se perde à travers sa lecture ». Par conséquent, lorsque le lecteur arrive à la fin du rapport, il ne devrait pas être surpris du jugement rendu.

Lors de nos entrevues, après que chaque participant ait énuméré tous les éléments contribuant à générer un jugement satisfaisant, nous voulions connaître la pondération qu'il accordait à chacun d'eux, total devant arriver à 100%. Les résultats sont présentés ci-dessous.

4.2.2.2 Pondération des éléments contribuant à générer un jugement satisfaisant

Dans le but de discriminer davantage les éléments qui engendrent un rapport d'évaluation satisfaisant, nous avons demandé aux six participants de classer par priorité leurs réponses. La plupart ont déclaré cette tâche comme étant très difficile, car ils considéraient tous les éléments d'égale importance. Toutefois, lorsque nous avons insisté pour obtenir des valeurs différentes, certains ont donné les réponses suivantes :

Deux répondants soulignent le fait que « la cohérence est [leur] priorité » et qu'ils lui accorderaient un pointage de 100%. Ainsi, ils ne parviennent pas à accorder une pondération aux autres éléments énumérés parce que ces éléments contribuent tous de façon importante à la cohérence du rapport.

Un client mentionne que l'élément le plus important pour lui serait «... la méthodologie dans un sens où on a couvert l'ensemble des points de vue [...]. Est-ce qu'on a vu l'ensemble des parties finalement? »

Un autre accorde 60% à la méthodologie, car « si au départ l'évaluation se retrouve avec des lacunes sur ce plan, c'est tout le reste du rapport qui se retrouve avec des sérieux problèmes de crédibilité ».

En résumé, compte tenu du grand niveau d'accord entre les participants, nous pouvons donc dire que les éléments *cohérence* et *méthodologie* jouent, à eux seuls, un rôle primordial sur la façon dont le jugement sera perçu.

4.2.2.3 Conclusion de l'objectif 2 : identifier les éléments contribuant à générer un jugement satisfaisant

En conclusion, suite à l'analyse du discours des évaluateurs et des clients interviewés, nous avons relevé cinq éléments qui contribuent, selon eux, à générer un jugement satisfaisant. Seules la description du programme et la cohérence sont les éléments à ne pas avoir été nommés par tous les participants. Toutefois, cela ne signifie pas qu'ils nient leur contribution, mais bien qu'ils n'en ont pas fait mention dans le cadre de l'entrevue.

Nous remarquons que les trois premiers éléments, soit *la question et/ou les objectifs d'évaluation*, *la description du programme*, *la méthodologie* représentent les *assises* du jugement. Quant au quatrième élément, les *informations recueillies*, il sert d'*appui* direct au jugement. Finalement, la *cohérence*, dernier élément, est l'élément transversal qu'il est nécessaire de retrouver entre tous les éléments nommés précédemment pour que le rapport soit harmonieux. Ainsi, tous contribuent à rendre le jugement satisfaisant. La figure 4.1 illustre cette synthèse.

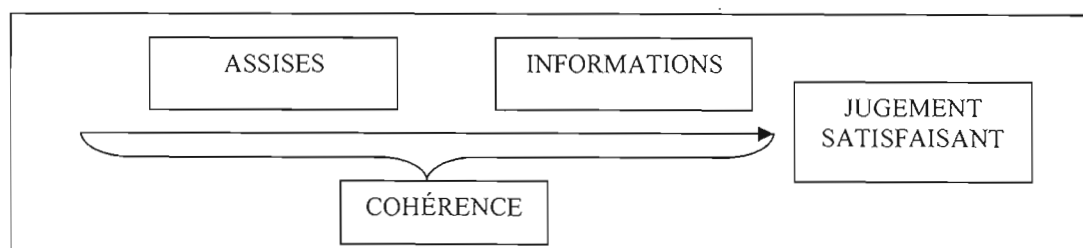


Figure 4.1 Synthèse des éléments qui contribuent à générer un jugement satisfaisant.

4.2.3 Objectif 3 : explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant

Comme nous l'avons déjà vu dans le cadre théorique de ce mémoire, Stake et Schwandt (2006) sont d'avis que tout évaluateur chevronné devrait recourir à la fois à la *qualité mesurée* et à la *qualité appréciée* pour appuyer son jugement évaluatif. Revoyons rapidement comment les auteurs distinguaient ces deux types de qualité. La qualité mesurée établit des évidences qui reposent sur des critères établis et opérationnels pour recueillir les données souhaitées, alors que la qualité appréciée se fonde sur la connaissance, l'expertise et l'intuition de l'évaluateur, mais aussi sur celles des personnes concernées par le programme afin de capter un autre type d'information, c'est-à-dire tout ce qui est non mesurable.

Lors de nos entrevues, spontanément, les répondants ont fait référence à la qualité appréciée, sans pour autant la nommer aussi clairement (voir l'élément « informations recueillies » de l'objectif 2). Toutefois, nous voulions approfondir cet aspect en lui accordant un traitement particulier. Alors, nous avons proposé aux participants les définitions de la qualité mesurée et de la qualité mesurée de Stake et Schwandt (2006) et nous leur avons demandé ce qu'ils en pensaient. Ensuite, nous les avons questionnés sur l'importance qu'ils accordaient à chacune de ces deux qualités et sur les éléments qui influencent, selon eux, la proportion du ratio de chacune des deux qualités. La synthèse de leurs réponses se trouve dans les deux prochaines sections.

4.2.3.1 L'importance de la qualité mesurée et de la qualité appréciée

Lorsque nous avons demandé aux participants s'ils considéraient important et pertinent qu'un jugement évaluatif soit appuyé par des arguments provenant à la fois de la qualité mesurée et de la qualité appréciée, tous ont répondu par l'affirmative. Selon eux, les deux types de qualités rentrent en ligne de compte dans le jugement parce qu'elles représentent les piliers du jugement. Cependant, quand nous les avons questionnés sur la façon de les utiliser,

leurs réponses variaient légèrement. Tandis qu'un participant insistait sur le fait que ces deux qualités devaient fonctionner en synergie plutôt qu'en vase clos, un autre disait qu'il se pouvait très bien qu'elles se fassent en parallèle, en série, ou de façon intercalée. « C'est comme si l'évaluateur utilisait deux paires de lunettes différentes pour observer et répondre aux mêmes questions d'évaluation », a mentionné un troisième participant.

Peu importe ces petits différends, tous s'entendent sur le fait que recourir à la qualité mesurée et à la qualité appréciée enrichit l'évaluation, améliore la validité interne du rapport et permet à l'évaluateur d'aller plus loin dans son jugement. Un participant voit même une certaine forme de triangulation qui a pour effet d'augmenter la cohésion et la crédibilité du jugement porté sur le programme. « Recourir aux deux types de qualité donne l'impression que l'évaluateur a été méticuleux dans son travail » souligne un dernier répondant.

Idéalement, il faudrait que le rapport soit bien équilibré, c'est-à-dire qu'il faudrait retrouver un bon équilibre dans la proportion des deux qualités utilisées. D'ailleurs, un participant leur accorde exactement la même importance et souhaite donc retrouver la même proportion (50-50) de qualité mesurée et de qualité appréciée dans un rapport. Deux autres donnent plutôt la proportion 60-40 (qualité mesurée et qualité appréciée), mais insistent sur le fait que cette proportion n'est pas toujours possible. Elle peut varier selon l'objet ou la nature de l'évaluation.

Par exemple, la qualité appréciée pourrait se retrouver en plus faible quantité dans une évaluation où il faut mesurer les effets directs d'un programme, tels que le niveau d'atteinte des objectifs du programme. Effectivement, lorsqu'il s'agit de recueillir des traces écrites ou lorsqu'il existe des objectifs bien définis et structurés pour lesquels il y a des indicateurs clairs et réalistes, l'évaluateur recourt davantage à la qualité mesurée.

À l'inverse, la qualité appréciée est plus facile à utiliser lorsqu'il s'agit de mesurer les effets indirects d'un programme ou lorsque très peu d'informations et de faits sont disponibles. Toutefois, tous les participants interviewés reconnaissent qu'il serait inconcevable qu'un jugement se base uniquement sur la qualité appréciée. Un rapport comme

celui-là aurait du mal à être crédible. Même un évaluateur chevronné et reconnu dans le milieu ne pourrait rendre un jugement du genre. Un minimum de données et d'évidence est nécessaire pour générer un jugement satisfaisant.

Bien qu'un rapport n'utilisant que de la qualité mesurée serait plus acceptable, les participants croient que l'évaluateur qui agirait ainsi finirait par paralyser l'exercice d'évaluation. « Il n'arrivera pas à saisir totalement le sens et l'essence de ce qui est là », mentionne un participant. Un pan de l'intervention peut échapper à l'évaluateur au moment où il choisit ses méthodes de collecte de données et s'il fait une erreur à cette étape, une partie du reste de son évaluation sera faussée. Un second participant renchérit:

« I have seen some evaluations that are technically great. They've got the methods, they've collected the data, and there's just no heart ! There's just no story, no sense that the evaluators really understood the context of the program or that they are understanding the environment of specific aspects of the delivery. It's just Cold! So, when it's like that, you're kind of left feeling... did you really get the whole story? »

Plus précisément, la qualité mesurée est la base et le fondement de l'évaluation. Tous les participants s'entendent pour dire qu'il est relativement facile d'établir un jugement en s'appuyant sur la qualité mesurée, soit sur des données et des évidences reposant sur des critères et des seuils de performance établis à priori. Généralement, ce type de qualité est solide, claire, facile et très explicite, ce qui fait que les évaluateurs choisissent souvent d'y recourir pour se rassurer. Du moment que «...l'évaluateur n'intervient pas comme un simple technicien qui traite une commande sans réfléchir», elle a pour effet de rassurer le client, précise un des participants.

La qualité appréciée réfère aux différentes perceptions des détenteurs d'enjeux. Ainsi, un jugement est plus crédible et plus facilement acceptable si ces derniers ont l'impression que leurs opinions et points de vue ont été pris en compte dans le jugement rendu. Généralement, il s'agit d'informations qui étaient impossibles à observer à l'aide des outils de collecte parce qu'elles n'étaient pas prévues (ex. : informations circonstanciées dans le discours d'un client) ou parce qu'il n'existait pas de critères et d'indicateurs mesurables de l'objet d'évaluation (ex. : désorganisation sociale). Grâce à son expertise, son instinct professionnel

et ses impressions, l'évaluateur peut relever ces éléments inexplorés ou difficiles à camper. « C'est comme si l'évaluateur possédait de longues antennes très aiguisées et très sensibles » qui lui permettent d'aller chercher ces nouveaux éléments, déclare un participant. Ceux-ci peuvent confirmer ou contredire les éléments recueillis à l'aide de la qualité mesurée, mais souvent, ils permettent de poser un regard différent, une vision critique sur les données amassées préalablement. Ultimement, ils permettent d'alimenter, de nuancer et d'enrichir le jugement final. Ce sont pour toutes ces raisons que l'évaluateur ne doit pas se restreindre à ce qui était prévu, mais plutôt être ouvert et attentif à tout ce qui émerge.

Bien que tous les participants s'entendent sur l'importance de la qualité appréciée, plusieurs trouvent difficile de lui donner un maximum de sens et de la rendre aussi rigoureuse que la qualité mesurée afin qu'elle soit reconnue comme valide. Elle n'est pas évidente à insérer dans un rapport, surtout quand elle contredit la qualité mesurée. Le pouvoir d'argumenter n'est pas aussi convaincant parce que les éléments soulevés ne sont pas évidents à documenter, à traduire et à synthétiser. Il faut que l'évaluateur soit capable de démontrer que le jugement qu'il porte est basé sur ce qu'il a perçu, d'une chronique qu'il a tenue, de ses observations non mesurées sur le terrain. L'évaluateur doit convaincre le client que le jugement qu'il porte ne dépend pas de ses valeurs personnelles, ni de son humeur de la journée.

Bref, pour les répondants, les éléments d'information venant de la qualité appréciée sont tout aussi importants que ceux venant de la qualité mesurée, c'est-à-dire qu'ils n'utilisent pas ces premiers pour illustrer le jugement rendu, mais qu'ils les traitent comme de l'information en soi. Ainsi, en combinant les deux qualités au même titre, l'évaluateur arrive à combler les écueils de chacune et s'assure d'une compréhension exhaustive du programme.

4.2.3.2 Les éléments influençant la proportion du ratio de la qualité mesurée et de la qualité appréciée

Notre prochaine question d'entrevue consistait à demander aux évaluateurs et aux clients de l'évaluation quels étaient les éléments qui influencent, selon eux, la proportion du ratio qualité appréciée et qualité mesurée. Suite aux réponses obtenues, nous avons remarqué qu'il était possible de créer deux grandes catégories dans lesquelles chacun des éléments nommés pouvait être répertorié, soit les éléments qui sont hors du contrôle de l'évaluateur et ceux qui sont sous son contrôle.

Les éléments hors du contrôle de l'évaluateur

Il se dégage de l'analyse des données trois éléments sur lesquels l'évaluateur n'a pas de contrôle, mais qui exercent une influence sur la proportion du ratio qualité appréciée et qualité mesurée que l'évaluateur utilise. Ainsi, selon les perspectives des répondants interviewés, il s'agirait de la nature du mandat, de la nature du programme et de la taille du programme.

- La nature du mandat

La proportion de la qualité mesurée et de la qualité appréciée peut varier selon les questions d'évaluation, les objectifs d'évaluation ou la raison de la demande d'évaluation. À titre d'exemple, s'il s'agit d'un contexte de reddition de compte l'évaluateur tendra davantage vers la qualité mesurée pour établir son jugement. Au contraire, une des raisons pouvant faire pencher l'évaluateur vers la qualité appréciée est lorsqu'il fait face à des contraintes (monétaires, temporelles, du milieu, etc.) l'empêchant d'aller chercher toutes les données voulues.

- La nature du programme

Plus un programme est structuré, avec 1) des objectifs et des interventions clairs; 2) une population cible bien identifiée et; 3) des guides de formation et d'animation disponibles, plus l'évaluateur aura tendance à recourir à la qualité mesurée. Inversement, plus les éléments d'un programme sont complexes à saisir surtout au niveau des interventions et services, plus l'évaluateur tendra vers la qualité appréciée.

- La taille du programme

Certains programmes, notamment ceux du gouvernement fédéral, ont de multiples facettes: plusieurs départements comprenant des sous-programmes. Cela exige de nombreuses évaluations et oblige l'évaluateur à être très méthodique. L'évaluateur doit être en mesure de nager à travers les différents paliers (micro, méso ou macro) pour recueillir toutes les informations nécessaires. Évidemment, dans un tel contexte, la qualité mesurée se trouve être la méthode la plus efficace.

Les éléments sous le contrôle de l'évaluateur

Lors des entrevues, d'autres éléments ayant un impact sur la proportion du ratio qualité appréciée et qualité mesurée utilisée pour appuyer le jugement sont ressortis. Les répondants ont déclaré que la compétence, l'expérience professionnelle et la personnalité de l'évaluateur jouent également un rôle. Contrairement aux éléments nommés précédemment, ceux-ci sont sous le contrôle de l'évaluateur.

- La compétence de l'évaluateur

Un évaluateur compétent possède une bonne connaissance pratique ainsi qu'une certaine expertise du domaine à évaluer (ex. : la population visée par le programme). Ainsi, l'évaluateur possédant ces caractéristiques est davantage orienté vers le type de données à considérer. De plus, il « est capable de capter des éléments qui émergent en cours de route, puis de les assembler, de les trier, de les analyser et de faire du millage là-dessus », et ce, tout en s'assurant d'atteindre les objectifs fixés au départ. Ce qui revient à dire que plus un évaluateur est compétent, plus il est en mesure de baser son jugement sur la qualité appréciée.

- L'expérience de l'évaluateur

Un évaluateur expérimenté est un évaluateur qui a fait plusieurs évaluations dans lesquelles il a été exposé à toutes sortes de situations et de contextes. Il possède donc une grande référence qui lui permet de « voir la forêt plutôt que l'arbre ». Il a développé un instinct professionnel lui permettant d'anticiper l'allure qu'une évaluation peut prendre. Il est également plus à l'aise à se réajuster si un problème inattendu survient et ne se met pas autant de limites qu'un évaluateur junior.

L'évaluateur expérimenté a plus de facilité à aller chercher l'information pour ensuite être en mesure de bien comprendre l'environnement dans lequel se déroulent le programme, les objectifs et la clientèle visée par le programme. Cela est dû entre autres parce qu'il utilise les gens en place, leur vécu et leurs connaissances, mais aussi parce qu'il se réfère à ce qui s'est fait de semblable dans des recherches antérieures. Bref, il est capable de démontrer qu'il est allé au-delà de la simple méthode, procédure ou démarche d'évaluation comme le ferait un technicien.

Ainsi, de la même façon qu'un évaluateur compétent, plus l'évaluateur est expérimenté, plus il a de chance de recourir à la qualité appréciée pour teinter son jugement parce que,

normalement, au fil des ans, il a développé un sens critique, une ouverture d'esprit et une confiance en soi. Finalement, il est plus facilement capable de démontrer que le jugement qu'il porte est scientifiquement valide, et ce, même s'il repose en grande partie sur des dimensions appréciées.

Toutefois, les participants interviewés insistent sur le fait qu'il est possible que certains évaluateurs expérimentés n'aient jamais recours à la qualité appréciée pour fonder leur jugement.

- La personnalité de l'évaluateur

Selon les répondants, l'évaluateur passionné qui aime le programme et qui démontre de l'empathie envers les preneurs de décisions ou les clients du programme aura plus tendance à recourir à la qualité appréciée pour compléter ou nuancer les données recueillies en qualité mesurée. Il ira probablement au-delà du minimum requis, tel que poser un regard critique sur les données amassées, s'il est conscient de tout l'impact que peut avoir son jugement sur l'orientation future du programme. Sachant que les gens impliqués dans le programme ont une perception plus appréciée du programme, un évaluateur ouvert d'esprit, à l'écoute et sensible à son environnement n'hésitera pas à tenir compte de leurs points de vue. À l'opposé, un évaluateur plus paresseux ou mercantile risque de n'utiliser que la qualité mesurée pour argumenter son jugement.

4.2.3.3 Conclusion de l'objectif 3 : explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant.

En résumé, tous les répondants s'entendent pour dire que pour qu'un jugement soit reconnu comme satisfaisant, il doit obligatoirement s'appuyer sur la qualité mesurée et sur la qualité appréciée. De plus, ceux-ci notent six éléments qui influencent la proportion du ratio

des deux types de qualité utilisées par l'évaluateur pour argumenter son jugement évaluatif. La nature du mandat, la nature du programme et la taille du programme sont des éléments extérieurs à l'évaluateur. Plus ils sont complexes, plus l'évaluateur se tourne vers la qualité mesurée pour des raisons de sécurité et d'efficacité. La compétence, l'expérience et la personnalité de l'évaluateur font partie des éléments directement reliés à l'évaluateur lui-même. Grâce à ses connaissances dans le domaine à évaluer et à son instinct professionnel, mais surtout à sa bonne volonté, l'évaluateur est en mesure de capter des informations importantes qui n'étaient pas prévues, informations qui l'aideront à pondérer son jugement final. Ces trois derniers éléments réfèrent donc à la crédibilité de l'évaluateur. Plus le client accorde de la crédibilité à l'évaluateur, plus le jugement qui repose sur de la qualité appréciée sera accepté et reconnu.

4.2.4 Objectif 4 : explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant

Bien que les objectifs précédents traitaient de ce qu'il faut retrouver à l'intérieur d'un rapport pour obtenir un jugement satisfaisant, n'oublions pas que la présentation du rapport en lui-même a également un impact sur le niveau de satisfaction du jugement.

Tout d'abord, la présentation du rapport doit être adaptée à la clientèle qui le reçoit. De plus, si on retrouve des tableaux synthèse et que l'évaluateur accompagne son rapport d'une présentation orale, les chances de rendre un jugement satisfaisant sont optimisées.

- Un rapport adapté à la clientèle qui le reçoit

L'évaluation change et aujourd'hui les clients n'ont plus le temps de lire des rapports trop épais. Ainsi, il semble qu'un rapport de dix à quinze pages accompagné d'un sommaire exécutif de deux ou trois pages est amplement suffisant en ce qui concerne les documents

remis aux clients de l'évaluation. Chacune des sections du rapport doit être synthétisée au maximum et aller directement à l'essentiel.

De plus, généralement les clients se plaignent de trouver les rapports beaucoup trop complexes. Ils souhaitent des résultats présentés simplement pour lesquels il n'est pas nécessaire d'être statisticien pour les comprendre. Le rapport doit être écrit dans un vocabulaire accessible de façon à ce qu'il puisse être lu par le plus de gens possible. Ainsi, il sera plus facile pour eux de se réapproprier l'information contenue dans le rapport pour pouvoir la réinvestir dans leur programme.

- Un rapport comprenant des tableaux

Le rapport doit inclure des tableaux synthèse qui présentent l'information clairement et qui aident le client à trouver rapidement ce qu'il cherche. Ces tableaux doivent comprendre chacune des questions d'évaluation, les indicateurs utilisés et les résultats de recherche qui y sont associés. Les tableaux évitent au lecteur de se noyer dans l'information ou de s'accrocher dans des détails. En un coup d'œil, il est possible d'avoir toute l'information importante de façon épurée et structurée.

- Un rapport accompagné d'une présentation orale

Le rapport doit être accompagné d'une présentation orale interactive auprès des clients, telle qu'une présentation multimédia. Lors de cette présentation, l'évaluateur en profite pour soulever les éléments moins formels qu'il n'a pas voulu intégrer dans le rapport. Il justifie certains points. Puis, il vérifie la compréhension des résultats, du jugement et des recommandations par les clients. Il les aide à se réapproprier l'information, à intégrer les résultats et à choisir les actions qui seront prises pour assurer la suite de l'évaluation. Finalement, il y a un partage d'opinions et de réactions entre toutes les personnes impliquées.

Ainsi, la présentation orale a pour but d'augmenter les chances d'utilisation du rapport d'évaluation et de ses recommandations. Comme le mentionne une personne que nous avons interviewée, « c'est plus de la façon dont l'évaluateur parle de l'évaluation qu'à travers le rapport qu'on sent si le jugement a de l'allure, s'il nous convainc. »

4.2.5 Synthèse de la présentation des résultats

Notre analyse des données nous a permis de dégager certains constats qui, s'ils ne font pas toujours l'unanimité, forment un consensus chez les répondants, évaluateurs comme clients de l'évaluation. Leurs réflexions nous ont permis d'établir un portrait de ce que constitue un jugement satisfaisant aux yeux des clients de l'évaluation ainsi que des procédures pour arriver à le produire.

À partir de nos objectifs de recherche et de l'analyse des données recueillies en entrevue, nous proposons la figure synthèse située à la page suivante (voir fig. 4.2), laquelle donne un rapide aperçu de nos résultats. Les cases en gris représentent des questions générales en lien avec nos objectifs de recherche, tandis que les cases en blanc constituent les réponses à ces questions, lesquelles sont tirées de l'analyse des données.

En résumé, selon les évaluateurs et les clients interviewés, un jugement satisfaisant est un jugement bien argumenté, explicite et utile. Pour y arriver, l'évaluateur doit :

- 1) Traiter de certains éléments dans son rapport d'évaluation. Il veillera à choisir une question d'évaluation et/ou des objectifs en lien avec les intérêts des détenteurs d'enjeux. Il fera également une description détaillée du programme pour démontrer qu'il en a bien compris les composantes et l'environnement dans lequel il se trouve. Puis, afin de donner une légitimité scientifique à son jugement, il veillera à ce que son rapport contienne une section « méthodologie ». Aussi, il présentera les informations sur lesquelles son jugement repose. Finalement, il s'assurera que tous ces éléments soient cohérents entre eux.

2) Recourir autant à la qualité mesurée qu'à la qualité appréciée pour argumenter et nuancer son jugement. Toutefois, certains éléments peuvent influencer la proportion accordée à chacun des deux types de qualité. Ainsi, nous retrouvons des éléments hors du contrôle de l'évaluateur (caractéristiques du programme et la nature du mandat) et d'autres sous son contrôle (sa compétence, son expertise et sa personnalité).

3) Présenter l'information contenue dans le rapport de façon à ce qu'elle soit adaptée aux lecteurs potentiels, synthétisée sous la forme de tableaux et accompagnée d'une présentation orale.

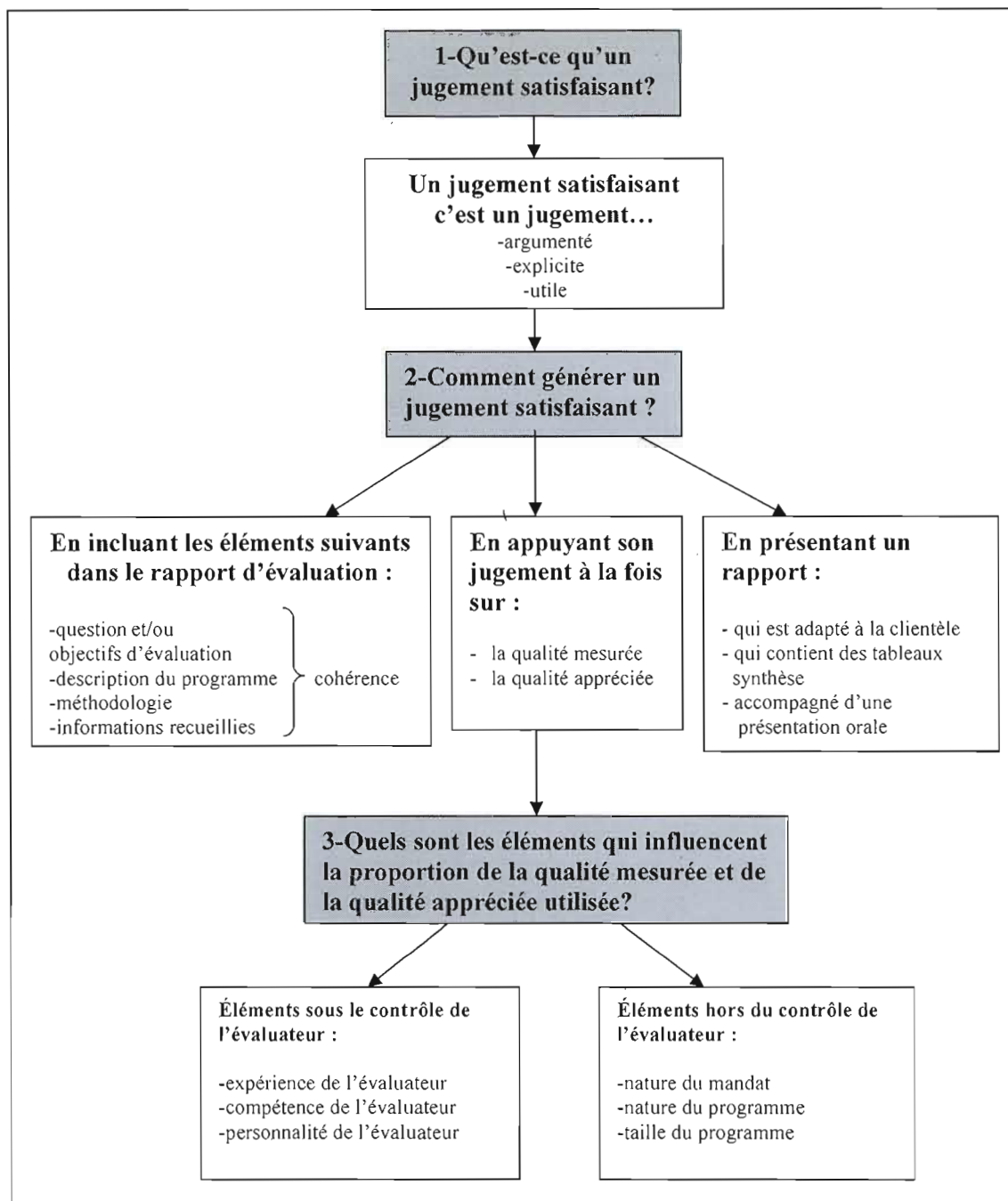


Figure 4.2 Synthèse des résultats de nos objectifs de recherche sur les éléments et les procédures qui contribuent à générer un jugement satisfaisant aux yeux des clients de l'évaluation.

CHAPITRE V

DISCUSSION

Ce cinquième et dernier chapitre a pour but de discuter des résultats rapportés dans le chapitre précédent. Nous ferons ressortir les points culminants qui se sont dégagés de l'analyse des résultats des entrevues individuelles portant sur le jugement évaluatif satisfaisant. À partir de là, nous établirons des liens avec les études recensées dans notre cadre théorique. Puis, nous terminerons ce chapitre par les questions qui restent en suspens.

5.1 Les résultats les plus révélateurs dégagés de l'analyse des données

Certains résultats méritent d'être discutés plus particulièrement dans cette section car ils rejoignent les perspectives de théoriciens. Ainsi, il ressort des résultats de notre recherche qu'évaluateurs chevronnés, clients de l'évaluation et théoriciens s'entendent sur : 1) *les critères d'un jugement satisfaisant*; 2) *la teneur politique de l'évaluation et l'implication des détenteurs d'enjeux*; 3) *l'importance de considérer tant la mesure que l'appréciation*. Ces trois points sont maintenant discutés ci-dessous.

1) Les critères d'un jugement satisfaisant

Lors des entrevues, les répondants ont mentionné qu'un jugement doit être *explicite* pour qu'ils le considèrent comme étant satisfaisant. Ce constat va dans le même sens que ce que Van der Maren (1995) souligne à propos des stratégies de la recherche spéculative. Selon cet auteur, les présupposés et les concepts utilisés par l'évaluateur doivent être préalablement clarifiés. La rédaction du rapport, et du jugement par le fait même, doit être faite de façon à ce que tous les éventuels lecteurs soient en mesure de comprendre facilement la terminologie utilisée.

De plus, nous avons vu que les praticiens interviewés considéraient l'*utilité du jugement* comme un second critère pour pouvoir qualifier un jugement de satisfaisant. Ce critère rejoint la perspective d'Alkin et Christie (2004) lorsqu'ils font allusion à la branche d'arbre sur laquelle se retrouvent les modèles théoriques reliés à l'*utilisation* de l'évaluation. Shadish, Cook et Leviton (1991) avaient également soulevé cet élément dans l'une des cinq composantes de leur métathéorie. Ainsi, il semble que théoriciens et praticiens s'entendent pour dire qu'il est primordial que les résultats de l'évaluation soient utiles et récupérables par les détenteurs d'enjeux pour les prises de décision quant aux orientations futures du programme.

Un jugement qui est *argumenté* est le dernier critère relevé par les répondants en entrevues. Nous avons vu que les praticiens se sont exprimés de façon spontanée, sans pour autant parler dans des termes aussi définis, de l'importance d'appuyer le jugement par de l'information provenant de deux sources : de données planifiées, recueillies à l'aide d'instruments de mesure, et d'informations pertinentes qui émergent au cours de l'évaluation. Il est donc facilement possible de relier ces définitions aux concepts de qualité mesurée et de qualité appréciée que Stake et Schwandt (2006) ont proposé dans leur article *On discerning quality in evaluation*. Ces auteurs vont plus loin dans leur définition de la qualité appréciée en mentionnant que celle-ci provient de toutes les expériences, les opinions et les connaissances des détenteurs d'enjeux. Étant donné que ces derniers n'ont pas les mêmes perspectives que celles de l'évaluateur, il est alors plus facile de tracer un portrait juste du programme en question.

Somme toute, les trois critères soulevés par les praticiens interviewés sont appuyés par des théoriciens, ce qui porte à croire qu'il s'agit réellement de critères essentiels pour générer un jugement satisfaisant. Toutefois, en travaillant en solitaire, l'évaluateur ne peut y arriver. Que ce soit pour s'assurer que les résultats et recommandations de l'évaluation soient utilisés ou que ce soit pour recueillir des informations basées sur la qualité appréciée, il est nécessaire que l'évaluateur sollicite l'implication des détenteurs d'enjeux, et ce, tout au long de la démarche évaluative. Ceci nous amène à traiter du point suivant.

2) La teneur politique de l'évaluation et l'implication des détenteurs d'enjeux

Nous avons déjà vu que les critiques adressées à l'évaluation de programme concernent son incapacité à générer un jugement qui satisfait les clients de l'évaluation. Les répondants interrogés en entrevue ont soulevé le stress et les tensions qui surviennent souvent entre les évaluateurs et les clients de l'évaluation lorsque vient le temps de faire approuver le jugement évaluatif. Les détenteurs d'enjeux veulent de plus en plus avoir un contrôle de l'information générée, ce qui entraîne une politisation de l'évaluation (House, 2003 : voir House et Howe, 1999; Stake et Schwandt, 2006). C'est pourquoi le jugement rendu doit non seulement s'appuyer sur des données qui ont été recueillies à partir d'une méthodologie solide (qualité mesurée), mais aussi sur les différentes perspectives, opinions et connaissances des détenteurs d'enjeux (qualité appréciée). Cela revient donc à dire que l'évaluateur doit instaurer entre les différentes parties concernées par l'évaluation un processus de démocratisation dans lequel il adopte un rôle de négociateur pour amener des compromis. D'ailleurs, dans son article *Evaluation, democracy, and social change*, Greene (2006) souligne que l'évaluation démocratique fournit des balises essentielles à la pratique pour assurer une crédibilité à l'évaluation, critère important aux yeux de tous.

Selon les répondants interrogés en entrevue, le programme s'inscrit dans une réalité sociale où se retrouve un ensemble de croyances et de valeurs que l'évaluateur ne peut pas toujours mesurer à l'aide d'instruments de mesure. Plus le jugement reflète la réalité, plus il devient difficilement contestable. Il est donc nécessaire que l'évaluateur aille chercher cette information en dialoguant avec les détenteurs d'enjeux. En sollicitant la participation des différents détenteurs d'enjeux, l'évaluateur recueille diverses perspectives aussi bien au niveau des besoins que des attentes, des opinions ou des points de vue. Encore une fois, Greene (2006) note une tendance de plus en plus forte à faire participer les détenteurs d'enjeux au processus évaluatif. En effet, les détenteurs d'enjeux reconnaissent qu'une évaluation est satisfaisante si elle réussit à cerner l'ensemble des enjeux essentiels d'un programme. Pour y arriver, il est nécessaire de recueillir les différents points de vue des détenteurs d'enjeux. Les propos de Weiss (1981) vont dans le même sens. Selon cette auteure, chaque programme est unique puisqu'il évolue à l'intérieur d'un contexte qui lui est

spécifique, sans compter les dimensions relationnelles qui entrent en interaction avec les dimensions techniques du programme. Ainsi, le dialogue avec les détenteurs d'enjeux est nécessaire. Il permet à l'évaluateur de saisir le programme en entier ainsi que l'environnement dans lequel il s'insère. House et Howe (1999) ainsi que Stake et Schwandt (2006) ont aussi mentionné l'importance d'instaurer une bonne communication entre les détenteurs d'enjeux pour que l'évaluateur soit en mesure de mieux comprendre la complexe réalité du programme. L'évaluateur doit favoriser les échanges de perspectives sur les multiples réalités et négocier jusqu'à l'obtention d'un consensus. Shadish, Cook et Leviton (1991) ainsi qu'Alkin et Christie (2004) rejoignent aussi cette perspective puisqu'ils ont traité de l'importance des différentes valeurs pour appuyer le jugement.

Théoriciens et praticiens reconnaissent donc qu'il est nécessaire d'adopter une vision démocratique de l'évaluation dans laquelle l'évaluateur prend en considération les différentes perspectives des détenteurs d'enjeux (Stufflebeam, 2000; House et Howe, 1999). Mais l'évaluateur « traditionnel » est-il prêt à troquer son rôle traditionnel pour celui de négociateur? Est-il prêt à laisser une aussi grande place aux détenteurs d'enjeux?

3) L'importance de considérer tant la mesure que l'appréciation

Les témoignages des personnes interviewées vont dans le même sens que les propos tenus par Stake et Schwandt (2006) à l'effet qu'on devrait toujours retrouver dans toute évaluation des arguments s'appuyant sur la qualité mesurée et d'autres sur la qualité appréciée, et ce, dans des proportions variant selon l'objet d'étude et le contexte du programme. Ainsi, l'évaluateur qui recourt aux deux types de qualité de façon complémentaire pour élaborer son jugement parvient à donner un maximum de crédibilité à ce dernier. L'important est que l'évaluateur précise bien sur quelles sources d'information repose son jugement.

Les personnes interviewées semblent conscientes de l'importance de leur rôle sur les suites du programme à l'étude et c'est pour cette raison qu'elles ne veulent pas se fier

uniquement à la qualité mesurée. Selon elles, bien que cette dernière soit importante, elle n'est toutefois pas suffisante. Certains reconnaissent que le jugement de l'évaluateur s'avère aussi complexe que celui d'un médecin ou d'un juge, lesquels ne se fient pas uniquement aux données recueillies, mais également au discours du malade, du plaignant ou de leurs proches avant d'établir un diagnostic ou un verdict. Dans le même ordre d'idées, Stake et Schwandt (2006) sont d'avis que plus de reconnaissance devrait être accordée à la qualité appréciée en évaluation de programme, laquelle permet à l'évaluateur non pas d'agir comme simple technicien, mais comme un professionnel qui se sert de son expertise et de celle des autres pour juger une situation.

5.2 Les questions qui restent en suspens

Il ressort des entrevues que si on a bien cerné les aspects de légitimité scientifique, soit la qualité mesurée, les répondants se sentent mal à l'aise en ce qui concerne le traitement des informations qui s'appuie sur la qualité appréciée. Les répondants ont insisté sur le fait qu'il n'est pas toujours facile de porter un jugement qui s'appuie sur la qualité appréciée parce que ce type d'information n'est pas facilement reconnu comme rigoureux. Il est donc important de documenter solidement la qualité appréciée. Bien que ce type de démarche génère un nombre important d'informations, l'évaluateur doit être en mesure de dégager les éléments superflus des éléments pertinents. À ce propos, Datta (2006) souligne que le défi qui reste est de savoir agencer les informations provenant des deux types de qualité afin de générer un jugement qui satisfasse les clients de l'évaluation.

House et Howe (1999) sont du même avis. Selon eux, la conception actuelle de l'évaluation de programme est encore très limitée en ce qui concerne la reconnaissance accordée à la qualité appréciée dans la démarche évaluative. Il existe encore beaucoup trop de clivage entre la qualité mesurée et la qualité appréciée et il faudrait qu'un jour, la pratique en évaluation de programme accorde ses lettres de noblesse autant à un paradigme qu'à un autre et qu'elle arrête de voir les deux types de qualité comme des entités opposées. Au contraire, il

faudrait enfin réaliser qu'il n'est pas possible de résumer une entité d'une seule façon. Mais comment changer cette vision des choses? Bien que les évaluateurs chevronnés et les clients interviewés reconnaissent l'importance de recourir à la qualité appréciée et de prendre en considération les différents points de vue des détenteurs d'enjeux pour appuyer le jugement, certaines études montrent que ce n'est pas cela qui se passe dans la pratique. Toutlemonde (2005) a notamment remarqué la difficulté qu'ont les évaluateurs à collaborer avec les détenteurs d'enjeux. Pourquoi? Serait-ce un manque d'expérience, de compétence chez les évaluateurs? Est-ce qu'une formation obligatoire aiderait les évaluateurs à mieux recueillir et utiliser l'information générée par la qualité appréciée? Nadeau (1988) ainsi que Donaldson et Lipsey (2006) croient qu'une certification obligatoire serait peut-être une solution.

En résumé, nous croyons que nos résultats de recherche peuvent constituer des pistes d'explication intéressantes concernant les difficultés auxquelles fait face l'évaluation de programme. L'évaluation, comme tout domaine, évolue à travers le temps. Nous avons vu que les instances politiques souhaitent prendre de plus en plus de place dans la démarche évaluative et il semble que, selon nos résultats, la qualité appréciée pourrait être une piste de solution. Dans ce sens, il serait intéressant d'investiguer davantage sur la qualité appréciée. De plus, plusieurs auteurs, dont Chelimsky (2006), parlent de jugement « crédible », mais ne le définissent pas. Cette notion est similaire à celle de jugement « satisfaisant » puisque logiquement, un jugement qui est crédible est par conséquent satisfaisant. Nous croyons donc que nos résultats de recherche peuvent fournir des pistes pour opérationnaliser ce terme de crédibilité.

CONCLUSION

Les nombreuses critiques adressées à l'évaluation de programme remettent en cause la crédibilité de sa fonction première, soit générer un jugement capable d'aider au processus de prise de décision des activités futures du programme. La présente recherche exploratoire visait à mieux comprendre la production d'un jugement satisfaisant aux yeux des clients de l'évaluation. Elle s'appuie sur des informations recueillies auprès des personnes directement concernées par l'évaluation, soit des évaluateurs chevronnés et des clients. L'analyse des données recueillies lors des entrevues semi-structurées a permis de valider plusieurs réflexions émises par des théoriciens.

Tout d'abord, en ce qui concerne notre premier objectif de recherche, *établir ce qu'est un jugement satisfaisant*, les répondants ont adopté une position assez consensuelle. Pour qu'un jugement soit considéré comme satisfaisant à leurs yeux, il doit répondre à trois critères, soit être solidement argumenté, utile à la prise de décision des grandes orientations du programme et explicite.

Ensuite, les résultats obtenus au deuxième objectif, *identifier les éléments contribuant à générer un jugement satisfaisant*, révèlent que, pour y arriver, cinq éléments doivent être traités à l'intérieur du rapport. Ainsi, l'évaluateur doit veiller à choisir une question d'évaluation et/ou des objectifs en lien avec les intérêts des détenteurs d'enjeux, faire une description détaillée du programme et de la méthodologie utilisée, présenter les informations sur lesquelles son jugement repose et finalement, s'assurer que tous ces éléments soient cohérents entre eux.

Aussi, les résultats du troisième objectif de recherche, *explorer la contribution respective de la qualité mesurée et de la qualité appréciée dans l'élaboration du jugement satisfaisant*, indiquent que tous les répondants sont d'avis pour dire qu'un jugement satisfaisant doit obligatoirement s'appuyer sur la qualité mesurée et sur la qualité appréciée, mais dans des proportions qui peuvent varier. À ce propos, six éléments influençant la proportion du ratio

des deux types de qualité ont été relevés, soit des éléments sous le contrôle de l'évaluateur et d'autres hors de son contrôle, telles les caractéristiques du programme en question.

Enfin, les résultats du dernier objectif de recherche, *explorer la façon de présenter les éléments permettant de générer un jugement satisfaisant*, semblent démontrer que l'information contenue dans le rapport doit être adaptée aux lecteurs, synthétisée sous la forme de tableaux et accompagnée d'une présentation orale.

Somme toute, nos résultats nous ont permis de mieux saisir une partie du processus expliquant ce que constitue un jugement crédible ainsi que les conditions requises pour le produire. Nos résultats tendent à démontrer que les perceptions des répondants rejoignent celles des auteurs Stake et Schwandt (2006). Ainsi, la pertinence des deux espaces pour générer un jugement qui satisfait les clients serait reconnue tant par les praticiens que par les théoriciens.

Bien que nos résultats soient basés sur un petit échantillon, ils démontrent la pertinence de poursuivre les recherches sur ce qui constitue un jugement satisfaisant aux yeux des clients et des conditions requises pour y parvenir. Les constats émis dans cette étude peuvent servir de tremplin pour des recherches ultérieures. Par exemple, la figure synthèse proposée à la toute fin de la discussion pourrait éventuellement être reprise et approfondie. Également, il est important de mieux documenter la notion de qualité appréciée afin qu'un jour l'évaluation de programme lui accorde ses lettres de noblesse. Ultimement, nous espérons que nos pistes d'explication déboucheront sur des pistes de solution qui permettront d'améliorer la pratique en évaluation de programme, et qui, plus spécifiquement, aideront les praticiens à mieux soutenir leur argumentation afin que les recommandations émises soient utilisées dans les milieux.

APPENDICE A

QUESTIONNAIRE D'ENTREVUE

CONTENU (présence et pondération des éléments)

1. Si vous vous référez à l'ensemble des rapports d'évaluation (que vous avez rédigés ou dont vous avez pris connaissance) :
 - a. Sur quoi vous appuyez-vous pour établir qu'un rapport d'évaluation est satisfaisant?
 - b. Pouvez-vous pondérer les éléments que vous venez d'identifier en leur accordant un pourcentage relatif (le total devrait être de 100%) afin que nous ayons une vue d'ensemble sur leur importance respective?
2. En ce qui concerne plus particulièrement le jugement émis dans le cadre d'un rapport d'évaluation:
 - a. Quelle place lui accordez-vous (très importante? assez importante? peu importante? aucunement importante? vous êtes indifférent-e)? Pouvez-vous justifier votre réponse?
 - b. Quelle forme devrait-il prendre?
 - c. Sur quoi devrait-il s'appuyer?
 - d. Finalement, qu'est-ce qui fait que vous « achetez » un jugement?
3. *L'interviewer explique ce que les auteurs (Stake et Schwandt, 2006) entendent par qualité mesurée et qualité appréciée (définition fournie) ainsi que leur rôle dans la construction du jugement fondé. Dans cette perspective :*
 - a. Que comprenez-vous de ces définitions?
 - b. Considérez-vous cette distinction pertinente? Intéressante?
 - c. Selon vous, ces deux types de qualité devraient-ils toujours être présents dans une démarche évaluative? Et devrait-on leur accorder la même importance?

- d. Selon votre expérience, dans quelle mesure les facteurs suivants peuvent influencer l'utilisation des deux types de qualité:
- la nature du mandat?
 - l'expérience de l'évaluateur?
 - la taille du programme?
 - la marge de manœuvre laissée à l'évaluateur?
 - autres?
- e. Dans un autre ordre d'idées, avez-vous l'impression qu'il est requis, dans le cadre de toute évaluation, d'avoir recours à des 1) *critères*; 2) *seuils de performance*; pour être en mesure de développer un jugement ou une conclusion évaluative?
- f. Avez-vous l'impression que 1) *l'expérience*; 2) *l'instinct professionnel*; 3) *la connaissance pratique*; 4) *les impressions et perceptions*; 5) *les valeurs*; de l'évaluateur ont un impact sur le jugement ou sur la conclusion évaluative? Lequel a le plus d'impact selon vous?

CONTENANT (traitement et présentation des éléments)

4. Selon votre point de vue, quelle forme de présentation s'avère la plus adéquate en ce qui concerne les éléments qui viennent d'être discutés? Y a-t-il une façon de présenter l'information qui vous satisfait particulièrement?
5. Nous aimerions que vous imaginiez une situation dans laquelle vous n'auriez que deux minutes pour prendre connaissance d'un rapport d'évaluation. Que regarderiez-vous durant ce laps de temps?
6. Si vous aviez à présenter un rapport auprès d'une instance, quels éléments devraient être présents afin que vous vous sentiez tout à fait à l'aise parce que vous sentez le jugement fondé? Au contraire, quels éléments (présents ou absents) vous empêcheraient de vous sentir en confiance ou d'être à l'aise?
7. Finalement, y aurait-il un point qui vous apparaît important et que nous n'avons pas abordé dans le cadre de l'entrevue? Si oui, lequel?

APPENDICE B

FORMULAIRE D'INFORMATION ET DE CONSENTEMENT

Montréal, le _____

Formulaire d'information et de consentement

A) Renseignements aux participants

1. Objectifs de la recherche

La présente recherche se situe dans le cadre du programme de la maîtrise en éducation à l'Université du Québec à Montréal et se déroule sous la supervision de madame Marthe Hurteau, directrice de ce mémoire. L'objectif de cette étude est d'explorer les éléments qui contribuent à la production d'un jugement satisfaisant dans la pratique de l'évaluation de programme. Votre collaboration est sollicitée en tant que spécialiste de l'évaluation (ou de client) pour identifier des éléments d'information permettant de faire avancer les recherches scientifiques liées à ce domaine. À cet effet, une seule rencontre est prévue au cours de laquelle vous aurez à donner votre avis sur ce que représente, pour vous, un rapport d'évaluation satisfaisant. La durée prévue de la séance est d'environ une heure.

2. Confidentialité

Le présent projet a été soumis et accepté par le Comité institutionnel d'éthique de la recherche de l'UQAM. **L'anonymat et la confidentialité des informations recueillies seront préservés lors du traitement et de l'analyse des informations ainsi que lors de la diffusion des résultats de recherche.** Pour ce faire, un code vous sera attribué. **Sachez que, suite à votre consentement, l'entrevue sera enregistrée, mais que tous les enregistrements seront détruits à la fin de la recherche.** Les renseignements seront conservés dans un classeur sous clé situé dans un bureau fermé.

3. Droit de retrait

Votre participation à ce projet de recherche est entièrement volontaire, il sera donc possible de vous retirer à tout moment de la recherche, sans préjudice et sans devoir vous justifier. Si vous vous retirez de la recherche, les renseignements personnels recueillis vous concernant seront détruits.

4. Indemnité

Aucune indemnité n'est prévue pour votre participation à ce projet. Par contre, la diffusion des résultats pourra se faire en envoyant une demande par courrier électronique à : bossiroy.amelie@courrier.uqam.ca

B) Consentement

En tant que participant(e) à ce projet, je comprends que :

-Ma participation est volontaire et que je peux cesser d'y participer en tout temps sans avoir à fournir d'explications ou de raisons.

-Les informations que je divulguerais volontairement en entrevue seront enregistrées sur cassette audio.

-Toutes les informations ou tous les renseignements que je fournirai demeureront confidentiels et seront utilisés uniquement aux fins de la présente recherche.

-S'il advenait que des chercheurs autres que ma directrice et moi-même veuillent utiliser les informations ou renseignements recueillis, un nouveau consentement me serait demandé.

Je déclare avoir pris connaissance de ces informations, avoir obtenu les réponses à mes questions sur ma participation à la recherche et je consens librement à prendre part à ce projet de recherche.

Nom du participant (lettres moulées)

Signature du participant

Date

Signature de l'étudiante-chercheuse

Date

Signature de la directrice-chercheuse

Date

RÉFÉRENCES

- Alkin, M. C. 1990. *Debates on Evaluation*. Newbury Park: Sage Publications.
- Alkin, M. C., et C. A. Christie. 2004. «An Evaluation Theory Tree». In *Evaluation Roots: Tracing Theorists' Views and Influences*, sous la dir. de M. C. Alkin, p. 12-65. Thousand Oaks (Calif.): Sage Publications.
- Arens, S. A. 2005. *Understanding Practice Through Exemplary Evaluation Cases: Actes du congrès de la Société Canadienne d'Évaluation et de l'Association Américaine d'Évaluation* (Toronto, 24-30 octobre 2005).
- _____. 2006. «L'étude du raisonnement dans les pratiques évaluatives». *Mesure et évaluation en éducation*, vol. 29, no 3, p. 45-56.
- Blais, M., et S. Martineau. 2006. «L'analyse inductive générale: description d'une démarche visant à donner un sens à des données brutes». *Recherche qualitative*, vol.26, no 2, p.1-18.
- Borys, S., B. Gauthier, N. Kishchuk, et S. Roy. 2005. «Evaluation Practice in Canada National Survey». *Revue canadienne d'évaluation*, vol. 21, no 3. p. 32-45.
- Boutin, G. 1997. *L'entretien de recherche qualitatif*. Sainte-Foy: Presses de l'Université du Québec.
- Boutin, G., J.-P. Martinez, et Y. Montoya. 2006. *L'école actuelle face au changement : instruire, éduquer ou socialiser*. Coll. «Recherches enseignement, éducation et formation en espaces francophones», no 6. Québec: Presses de l'Université du Québec.
- Brousselle, A., F. Champagne et A.-P. Contandriopoulos. 2006. «Vers une réconciliation des théories et de la pratique de l'évaluation, perspectives d'avenir». *Mesure et évaluation en éducation*, vol. 29, no 3, p. 57-73.
- Chen, H.-T. 2005. *Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness*. Thousand Oaks (Calif.): Sage Publications.
- Christie, C. A. 2003. «What Guides Evaluation? A study of How Evaluation Practice Maps onto Evaluation Theory». *New Directions for Evaluation*, no 97, p. 7-36.
- Christie, C. A., et T. Azzam. 2005. «What Theorists Say They Do: A Brief Description of Theorists' Approaches». *New Directions for Evaluation*, no 106, p. 15-26.
- Davidson, E. J. 2005. *Evaluation Methodology Basics : the Nuts and Bolts of Sound Evaluation*. Thousand Oaks (Calif.): Sage Publications.

- Donaldson, S. I. et M. W. Lipsey. 2006. «Roles for Theory in Contemporary Evaluation Practice : Developing Practical Knowledge». In *The Sage Handbook of Evaluation*, sous la dir. de I. F. Shaw, J. C. Greene et M. M. Mark, p.56-75. Thousand Oaks (Calif.): Sage Publications.
- Dubois, N. et R. Marceau. 2005. «Un état des lieux théoriques de l'évaluation: une discipline à la remorque d'une révolution qui n'en finit pas». *La revue canadienne en évaluation de programme*, vol. 20, no 1, p. 1-36.
- Fine, A. H., C. E. Thayer, et A. T. Coghlan. 2000. «Program Evaluation Practice in the Nonprofit Sector». *Nonprofit Management and Leadership*, vol. 10, no 3, p. 331-339.
- Fink, A. 2004. *Evaluation Fundamentals Insights into the Outcomes, Effectiveness, and Quality of Health Programs*. Thousand Oaks (Calif.): Sage Publications.
- Fournier, D. 1995. «Establishing Evaluative Conclusions: A Distinction Between General and Working Logic». *New Directions for Evaluation*, no 68, p. 15-32.
- Fournier, D., et N. L. Smith. 1993. «Clarifying the Merits of Argument in Evaluation Practice». *Evaluation and Program Planning*, vol. 16, no 4, p. 315-323.
- Guba, E. G. 1972. «The Failure of Educational Evaluation». In *Evaluating Action Programs: Readings in Social Action and Education*, sous la dir. de C. H. Weiss, p.251-266, Boston: Allyn et Bacon Inc.
- Guba, E. G. et Y. S. Lincoln. 1981. *Effective Evaluation Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. San Francisco (Calif.): Jossey-Bass.
- _____. 1989. *Fourth Generation Evaluation*. Newbury Park (Calif.): Sage Publications.
- House, E. R. 1980. *Evaluating With Validity*. Thousand Oaks (Calif.): Sage Publications.
- House, E. R., et K. R. Howe. 1999. *Values in Evaluation and Social Research*. Thousand Oaks (Calif.): Sage Publications.
- Hurteau, M. 1991. *Les choix stratégiques en évaluation de programme*. Coll. «L'évaluation en action», no 1. Montréal: Centre d'évaluation de développement et d'intervention de Montréal.
- Hurteau, M., et S. Houle. 2005a. «Évaluation de programme et recherche évaluative: des activités distinctes». *Mesure et évaluation en éducation*, vol. 28, no 3, p. 83-95.
- Hurteau, M., et S. Houle. 2005b. *Identifying a Core Body of Knowledge for Evaluators: Actes du congrès annuel de l'Association Américaine d'Évaluation* (Toronto, octobre 2005).
- Hurteau, M., et S. Houle. 2006. *L'évaluation de programme: réflexion concernant la théorie de l'évaluation: Actes du 19^e colloque international de l'ADMEE-Europe* (Université du Luxembourg, septembre 2006).

- Hurteau, M., S. Houle, et S. Mongiat. (à paraître). «To What Extent Do Evaluative Practices Generate Legitimate Judgments?» Montréal: Université du Québec à Montréal
- Hurteau, M., G. Lachapelle, et S. Houle. 2006. «Comprendre les pratiques évaluatives afin de les améliorer: la modélisation du processus spécifique à l'évaluation de programme». *Mesure et évaluation en éducation*, vol. 29, no 3, p. 27-44.
- Hurteau, M., et M.-A. Nadeau. 1987. «Distinctions de base: évaluation de programme, recherche évaluative et recherche». *Revue canadienne de psycho-éducation*, vol. 16, no 2, p. 91-95.
- Johnson, R. L., M. J. Willeke, et D. J. Steiner. 1998. «Stakeholder Collaboration in the Design and Implementation of a Family Literacy Portfolio Assessment». *American Journal of Evaluation*, vol. 19, no 3, p. 339-353.
- Karsenti, T., et L. Savoie-Zajc. 2004. *La recherche en éducation: étapes et approches*. Sherbrooke: Éditions du CRP.
- King, J. A. 2003. «The Challenge of Studying Evaluation Theory». *New Directions for Evaluation*, no 97, p. 57-67.
- Landry, N., et R. Auger. 2007. *Stratégies de clarification conceptuelle*. Document RT_LABDORM300303. Laboratoire d'analyse de données et de formation en mesure et évaluation.
- Lessard-Hébert, M., G. Goyette, et G. Boutin. 1990. *Recherche qualitative fondements et pratiques*. Coll. «Éducation». Montréal : Agence d'Arc.
- Levin-Rozalis, M. 2003. «Evaluation and Research: Differences and Similarities». *La revue canadienne en évaluation de programme*, vol. 18, no 2, p. 1-31.
- Madaus, G. F., et T. Kellaghan. 2000. «Models, Metaphors, and Definitions in Evaluation». In *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, sous la dir. de D. L. Stufflebeam, G. F. Madaus et T. Kellagan, p. 19-31. Boston: Kluwer Academic Publishers.
- Madaus, G. F. et D. L. Stufflebeam. 2004. «Program Evaluation: A Historical Overview». In *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, sous la dir. de D. L. Stufflebeam, G. F. Madaus et T. Kellagan, p. 3-19. Boston: Kluwer Academic Publishers.
- Mark, M. M. 2008. «Building a Better Evidence Base for Evaluation Theory». In *The Fundamental Issues in Evaluation*, sous la dir. de N. L. Smith et P. R. Brandon, p. 111-134. New York: Guilford Press.
- Mathison, S. M. 2005. *Encyclopedia of Evaluation*. Thousand Oaks (Calif.): Sage Publications.
- McCarthy, T. A. 1973. «A Theory of Communicative Competence». *Philosophy of the Social Sciences*, no 3, p.135-156.

- _____. 1979. *A Communication and the Evolution of Society: Jürgen Habermas*. Toronto: Beacon.
- Nadeau, M.-A. 1988. *L'évaluation de programme théorie et pratique*. Sainte-Foy: Presses de l'Université Laval.
- Nevo, D. 2006. «Evaluation in Education». In *The SAGE Handbook of Evaluation*, sous la dir. de I. F. Shaw, J. C. Greene, et M. M Mark, p. 441-460. London: Sage Publications.
- Patton, M. Q. 1997. *Utilization Focused Evaluation: The New Century Text*. Thousand Oaks (Calif.): Sage Publications.
- _____. 2002. *Qualitative Research and Evaluation Methods*. Thousand Oaks (Calif.): Sage Publications.
- Redding, P. 1989. «“Habermas”, Theory of Argumentation». *Journal of Value Inquiry*, no 23, p. 15-32.
- Rog, D. J. 1995. «Reasoning in Evaluation: Challenges for the Practitioner». In *Reasoning in Evaluation: Inferential Links and Leaps*, sous la dir. de D.M. Fournier, p. 93-100. *New Directions for Evaluation*, no 68. San Francisco: Jossey-Bass.
- Rog, D. J., et D. Fournier. 1997. «Editors' notes». In *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods*, p. 1-13. *New Directions for Evaluation*, no 76. San Francisco: Jossey-Bass.
- Ryan, K. E. 1998. «Advantages and Challenges of Using Inclusive Evaluation Approaches in Evaluation Practice». *American Journal of Evaluation*, vol. 19, no 1, p. 101-122.
- Schwandt, T. A. 2007. *Educating for Intelligent Belief in Evaluation: Actes du congrès annuel de l'Association Américaine d'Évaluation* (Baltimore).
- Scriven, M. 1967. «The Methodology of Evaluation». In *Curriculum Evaluation: American Educational Research Association Monograph Series on Evaluation no 1*, sous la dir. de R.E. Stake. Chicago: Rand McNally.
- _____. 1980. *The Logic of Evaluation*. Inverness (Calif.): Edgepress.
- _____. 1990. «The Evaluation of Hardware and Software». *Studies in Educational Evaluation*, no 16, p. 3-40.
- _____. 1995. «The Logic of Evaluation and Evaluation Practice». *New Directions for Evaluation*, no 68, p. 49-70.
- _____. 2000. «Evaluation Ideologies». In *Evaluation Models: Viewpoints on Educational and Humain Services Evaluation*, sous la dir. de D. L. Stufflebeam, G. F. Madaus et T. Kellagan, p. 249-278. Boston: Kluwer Academic Publishers.

- Secrétariat du Conseil du Trésor du Canada. 2004. *Examen de la qualité des évaluations dans les ministères et les organismes*, Document interne, Gouvernement du Canada ([http : //www.tbs-sct.gc.ca/eval/pubs/rev-exam_f.asp](http://www.tbs-sct.gc.ca/eval/pubs/rev-exam_f.asp)).
- Shadish, W. E., T. D. Cook, et L. C. Leviton. 1991. *Foundations of Program Evaluation Theories of Practice*. Newbury Park: Sage Publications.
- Silvern, L. C. 1972. *System Engineering Applied to Training*. Houston: Gulf publishing Company.
- Smith, N. L. 1993. «Improving Evaluation Theory through the Empirical Study of Evaluation Practice». *Evaluation Practice*, no 14, p. 237-242.
- _____. 1995. «The Influence of Societal Games on the Methodology of Evaluative Inquiry». *New Directions for Evaluation*, no 68, p. 5-14.
- Stake, R. E. 1995. *The Art of Case Study Research*. Thousand Oaks (Calif.): Sage Publications.
- _____. 2001. «A Problematic Heading». *American Journal of Evaluation*, vol. 22, no 3, p. 349-354.
- _____. 2004. *Standards-based and Responsive Evaluation*. Thousand Oaks (Calif.): Sage Publications.
- Stake, R. E., C. Migotsky, R. Davis, E. J. Cisneros, G. DePaul, C. Dunbar, R. Famer, J. Feltovich, E. Johnson, B. Williams, M. Zurita, et I. Chaves. 1997. «The Evolving Syntheses of Program Value». *Evaluation Practice*, vol. 18, no 2, p. 89-103.
- Stake, R. E., et A. Schwandt. 2006. «On Discerning Quality in Evaluation». In *The Sage Handbook of Evaluation*, sous la dir. de I. F. Shaw, J. C. Greene, et M. M. Mark, p. 404-418. Thousand Oaks: Sage publications.
- Stevenson, J., et D. Thomas. 2006. «Intellectual Contexts». In *The Sage Handbook of Evaluation*, sous la dir. de I. F. Shaw, J. C. Greene, et M. M. Mark, p. 200-224. Thousand Oaks: Sage Publications.
- Stufflebeam, D. L. 2000. «Foundational Models for 21st Century Program Evaluation». In *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, sous la dir. de D. L. Stufflebeam, G. F. Madaus et T. Kellagan, p. 33-84. Boston: Kluwer Academic Publishers.
- Suchman, E. 1967. *Evaluative Research : Principles and Practice in Public Service and Social Action Program*. New York: Sage Publications.
- Taylor, P. W. 1961. *Normative Discourse*. Englewood Cliffs (N. J.): Prentice Hall.
- Thayer, C. E., et A. H. Fine. 2000. «Evaluation and Outcome Measurement in the Non-Profit Sector: Stakeholder Participation». *Evaluation and Program Planning*, vol. 23, no 1, p. 103-108.

- Thomas, D. R. 2006. «A General Inductive Approach for Analyzing Qualitative Evaluation Data». *American Journal of Evaluation*, vol. 27, p. 237-246.
- Toulemonde, J. 2005. *Appropriation des résultats de l'évaluation: Leçons de la pratique en Région Limousin: Acte du colloque de la Société Française de l'Évaluation* (mai 2005).
- Toulmin, S. E. 1993. *Les usages de l'argumentation*. Coll. «Interrogation philosophique». Paris: Presses universitaires de France.
- _____. 2001. *Return to Reason*. Cambridge (Mass.): Harvard University Press.
- Toulmin, S. E., R. D. Rieke, et A. Janik. 1984. *An Introduction to Reasoning*. New York: Macmillan.
- Unrau, Y. A. 2001. «Using Client Interviews to Illuminate Outcomes in Program Logic Models: A Case Example». *Evaluation and Program Planning*, vol. 24, no 4, p. 353-361.
- Valovirta, V. 2002. «Evaluation Utilization as Argumentation». *Evaluation*, vol. 8, no 1, p. 60-80.
- Van der Maren, J.-M. 1995. *Méthodes de recherche pour l'éducation*. Coll. «Éducation et formation». Montréal: Presses de l'Université de Montréal.
- Virtanen, P., et P. Uusikylä. 2002. *Exploring the Missing Links Between Causes and Effects: Conférence annuelle de la Société Européenne d'Évaluation* (Séville, octobre 2002).
- Weiss, C. H. 1983. «Ideology, Interests, and Information». In *Ethics, the Social Sciences and Policy Analysis*, sous la dir. de D. Callahan, et B. Jennings, p. 231-243. New York: Plenum Press.
- _____. 1998. *Evaluation Methods for Studying Programs and Policies*. Upper Saddle River (N. J.): Prentice-Hall.
- _____. 2004. «Rooting For Evaluation: A Cliff Notes Version of My Work». In *Evaluation Roots: Tracing Theorists' Views and Influences*. p. 153-168. Thousand Oaks, (Calif.): Sage Publications.